In this thesis, two problems are addressed.

The first problem is how designers of computerized artifacts come to understand what they should design. In order to improve their chances for solving the right problem, it is suggested to adopt a deductive theorizing protocol, based on falsification, as it is expressed in philosophy as well as in psychological research on problem solving and decision making.

It requires a shift in the role that designers take in communication with users; rather than considering users as objects of study, the designers' understanding is being scrutinized, by users.

To accomplish this, an explication of the designers' understanding is necessary to formulate: a Theory of Use, which should account for the assumptions that the designers make about the users and the use situation. It should also be focused on what the designers understand to be the qualities of a good solution. These assumptions and qualities should be formulated as statements that are falsifiable by the users. Through this, the designers can, in a systematic way, get feedback on how well they have understood what they are designing for. Furthermore, the Theory of Use becomes the requirements against which subsequent design hypotheses are evaluated.

The second problem is how to determine if a computerized artifact is appropriate for a given purpose. To guide such an evaluation, it is suggested to focus on the autonomy of the artifact's users. This is based on the philosophical and psychological conclusion that autonomy is necessary for human well-being, which is not explicitly taken into account in current design practice.

The proposed Theory of Use protocol also gives designers the autonomy they need for developing their understanding and expertise. It is intended to improve and guide communication with users, targeting the most important facet of design: the designers' understanding of what they are designing for. Moreover, the protocol puts users in control over what is developed for them, and lets them have the final say on how their lives are designed.

Keywords: design method, design philosophy, user-centered design, theorizing, falsification, interaction design, hypothesis, mental models, autonomy

Mikael Laaksoharju, Department of Information Technology, Division of Visual Information and Interaction, Box 337, Uppsala University, SE-751 05 Uppsala, Sweden.

© Mikael Laaksoharju 2014

ISSN 1104-2516
ISBN 978-91-554-8932-8
urn:nbn:se:uu:diva-221877 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-221877)

Printed by Elanders Sverige AB, 2014
Dedicated to you
who cannot bear taking responsibility
for what you know is wrong.
1. Preface

This thesis is about designing computerized artifacts. There are at least four ways to interpret it.

1. That I advocate designing for positive experiences.
2. That I want to increase user responsibility.
3. That I doubt the possibilities for solving the right problems in current system development practice.
4. That I fear technology.

All of these interpretations are correct, given specific qualifications.

1. I do think that it is important to aim for positive experiences, also for professional users. What I mean by positive experiences is, however, defined neither in hedonic nor ludic terms. I aim for more fundamental qualities of activity: the feeling of being autonomous and competent.
2. In order to address these more fundamental qualities, it is important that users are free to decide over relevant matters. Without freedom, there is no responsibility, and vice versa.
3. I do think that expert practitioners are good at identifying the right problems to address in systems development. However, expertise is created by reflecting over failure and success; by paying attention to what has worked well and what has not; and by developing the own understanding of the domain. It can be questioned whether current development practices support development of this expertise.
4. I will claim that humans should be in charge of technology, and that routines and regulations are part of the technological framework and part of what humans should be in charge of. It is questionable if that is the case today, but I believe that thoughtful design and development practice can reestablish this ideal. I wish to contribute to increased awareness of the responsibility that designers and developers have in shaping our future, and to increased awareness about what people at large should expect from technology.

1.1 The Author

I have studied computers and computing for quite some time now. I enjoy it and it has given me the degree Master of Science in Computer Science. On a
brief inspection, you will probably not notice much influence of this education in my thesis—most of the topics I refer to are more likely to be classified as philosophy or psychology, and what I write probably resides domain-wise between these areas.

My engineering schooling has taught me to break down problems into smaller problems and to pursue solutions to these. This approach may appear somewhat odd and unscientific to those scholars who are more accustomed to analyzing problems in depth. In a sense, what I have written is philosophy and psychology for engineers.

By hereby clarifying that I assume many engineers to have acquired an instinct to look for solutions to problems, and that this will affect how they make sense of the world, I hope to excuse the sometimes brief treatment of topics that, if sufficiently studied, in themselves could make up whole theses. I have been working by trying and failing, and trying again to produce practical tools to address questions that I consider important to ask. I do not claim that this practical manner of handling academic questions necessarily is a good approach—I may very well have missed important insights that would have improved my thinking—but it is solution-oriented and it produces something concrete. In best case, these concrete results support the thinking of others, and in worst case they stimulate your critical reflection.

This practitioner’s approach to research also means that I demand theories to have tool value and I evaluate understanding based on what it could possibly be useful for. This does not mean that I, for instance, think ethnography should be reduced to implications for design (see Dourish 2006 for a background), rather the contrary: I am afraid that ethnography loses its tool value if it is reduced to implications for design. Tool value is not synonymous with ease of use, a point which I will repeat several times in this thesis.

It does, however, mean that I think it is good to produce artifacts to demonstrate things. The leap from theory to practical application is sometimes trivialized within academia but it is not at all simple. In fact, the state of current technology should serve as ubiquitous proof that it is rather difficult.

The academic problem in my type of pragmatic approach, in which I introduce inexhaustive ideas from several different perspectives, is that I should still make claims that can be properly evaluated and not just passed on with a shrug and some remarks on insignificant details. Therefore, I have adopted a convention of stating clearly the assumptions that I am aware of making, as assertions. I try to refrain from using referenced work as arguments, to instead reconstruct the argument by logic and observation. Referenced work is from that point of view my empirical material. Those assumptions that I am not yet aware of, I anticipate with excitement to be informed about.

My main motivation for writing this thesis is to get feedback on my mental model of human-computer interaction, this wild-grown area of research—a mental model that I have managed to complicate even more for myself. I have been struggling against my old instincts and gone far outside my comfort zone
when developing the ideas that I put forth. Nevertheless, after my investigations, this is what I now believe is a sensible answer to why we still have to deal with badly designed artifacts. Only by exposing my thinking to others, will I be able to find out whether it really is.

1.2 Motivation

In my licentiate thesis, I posed the research question: “How should a computerized tool be designed so that it can stimulate decision makers to make morally good decisions?” (Laaksoharju 2010)

It is an idealistic question, which assumes that decision makers are willing to go to great lengths in order to make a moral decision. I did not fully believe that this would actually happen in practice. Morality can be learned, but it is the more difficult to teach and impose from outside. I will discuss this later. Nevertheless, if someone happened to be looking for computerized support for making that extra effort, she should at least be able to find something. That is how I used to think.

Over the years, as I have continued to think about what I proposed, the idea has matured into something more pretentious. Why settle with just those problems that appear to have a moral perspective? I was already claiming that there is a moral dimension to every problem so why should I not take the leap and claim that the process that I advocate is applicable to any problem?

In this thesis I propose and try to justify a shift of perspective for designing computing systems. It has implications for many aspects of computer use, not all of them necessarily attractive when viewed through a contemporary lens. Software and other artifacts may become uglier and more demanding to use—less “user-friendly”. Most importantly, however: if the individuals’ autonomy is considered the way I propose below, requirements to take responsibility for decisions will be amplified. Currently, automated systems are introduced to reduce human errors, not only in safety-critical systems. From my perspective, human errors are a necessary evil in a society where humans are treated as ends and where we want humans to act as ends. We should not sacrifice that which distinguishes us from machines, for qualities that are essentially nonhuman. Some human errors are, however, unnecessary: the ones that happen because humans have been deprived the possibilities to act responsibly. I hope this thesis is a small contribution to addressing that problem.

1.3 On Methods and Sources

Since this dissertation is for a technical degree, and presumably the reader is technically schooled, I feel obliged to comment a bit on the method that I have
been using in my research. Also the choice of sources probably requires some motivation, especially as they are rarely exhaustively referenced.

When it comes to methods, we want these to be predictable. Dictionary definitions usually bring up expressions like “systematic procedure”, “established”, and “organized”. Consequently, few established methods lead to completely unexpected results. They lead to answers to the questions that they were defined to investigate. In the area of HCI we have a plethora of methods to assess the usability of computer systems. These typically render the result that there are some usability problems in the inspected system. Sometimes more severe than anticipated and sometimes less so. They rarely, if ever, propose a completely new system as a result. Methods are tools and thus it is good if they have a predictable range of outcomes. When methods are used by insightful people, they can be very valuable by providing useful information. However, when used mindlessly, they can be disastrously misleading.

In this thesis I have not used such methods. To some it may occur like there is a fundamental lack of method in my work, and in one sense this is true, i.e., if a method is seen as something in line with the above definition. There is, however, a method in what I have done, although it is not guaranteed to be systematic and organized. It is the philosophical method of analysis and critical reflection. This method is not rigid in its procedure but it is extremely rigid in demanding justification for decisions. What matters is the arguments that you present, not how you found them. Arguments are scrutinized based on their assumptions and logical consistency. Despite me here appearing as a methodological heretic, I do understand, and partly sympathize, with the reasons for establishing certain methods as the norm for conducting research within a field. It makes research more efficient, and the researcher who is using them does not need to worry about epistemological arguments for how they fit research questions. Other researchers, who are expected to assess the work of their peer, will also quickly be able to understand what has been done and the quality of it. Many methods have also been refined over decades (if not centuries) to enable a rigor and validity that is difficult to achieve in a less conservative approach.

At the same time, if methods are given, there is little incentive for a researcher to form an independent understanding of their justification, with the effect that scientific rituals are applied also when they do not fit (Gigerenzer 2004). I did not choose voluntarily to devise my own methods, but when I scrutinized the ones that were offered to me, I failed to see how they fit my research. This may very well be a result of me asking the wrong questions, but I believe that within our field of research, Human-Computer Interaction, few of the methods that target creation of artifacts are epistemologically motivated. They are the result of best practice, often in alliance with business goals.
1.4 Disposition

In Part I, I give a preliminary introduction to the problems that I am to address, which some would rightfully regard as a philosophical question, and the pre-conditions for addressing this, i.e., the area of Human-Computer Interaction as an applied science. I also account for a minimal set of concepts that are relevant for the problem.

Part II gives an account of the theories that I base my reasoning on. I have made an effort to synthesize the different concepts into something that is coherent for my application.

Part III is a description of a protocol for theorizing about artifacts. I also give examples of how the theorizing process can be applied to address different types of problems. I have attempted to use distinctly different examples with varying levels of complexity to show the diversity of the process.

Part IV contains an English and a Swedish summary of the thesis, and a few suggestions for how to make use of it.

There is no specific part or chapter for discussions and you will notice that I engage in such wherever I find it necessary. This may violate some academic code, but, as will hopefully become clear, I think it is important to synthesize concepts, which necessitates occasional deviations from a strict reporting style. By this, I do not wish to give the impression that I think a discussion chapter is generally out of place. When, for instance, announcing the outcome of experiments, it is very important to maintain distinction between results and interpretation, as this allows others to interpret the data differently. In this thesis, however, in which I do not make my claims based on first-hand empirical findings, it seems unnecessary to make the distinction.
Contents

1 Preface ............................................................................................................. 7
  1.1 The Author ........................................................................................ 7
  1.2 Motivation ........................................................................................ 9
  1.3 On Methods and Sources ..................................................................... 9
  1.4 Disposition ......................................................................................... 11

Part I: Establishing the Question ..................................................................... 17

2 Introduction .................................................................................................... 19
  2.1 The Ultimate Decision Support System ............................................... 19
  2.2 Social Media as an Extension of Capabilities ....................................... 21

3 Science ......................................................................................................... 23
  3.1 Induction and Deduction ..................................................................... 23
  3.2 Abduction ........................................................................................... 25
  3.3 Logic .................................................................................................... 26
  3.4 Science and the Truth .......................................................................... 27
    3.4.1 The Truth .................................................................................... 28
    3.4.2 Objectivity .................................................................................. 29
    3.4.3 Values in Science ....................................................................... 30

4 Technology .................................................................................................. 31
  4.1 Artifacts ............................................................................................... 32
  4.2 Tools ..................................................................................................... 35
  4.3 Designers ............................................................................................. 39

5 Human-Computer Interaction ....................................................................... 41
  5.1 Birth and Focus of HCI ...................................................................... 42
    5.1.1 The State of HCI .......................................................................... 44
    5.1.2 The Future of HCI ....................................................................... 45
    5.1.3 The Human in HCI ...................................................................... 46
  5.2 Sketchpad—A Man-Machine Graphical Communication System .......... 48
  5.3 Ethical Interaction Design .................................................................... 51

Part II: Thinking Tools, a.k.a Theory ............................................................... 53

6 The Human Mind .......................................................................................... 55
  6.1 Mental Models ...................................................................................... 55
    6.1.1 The User’s Mental Model .............................................................. 58
6.1.2 The Designer’s Mental Model ........................................... 59
6.2 System I and System II Thinking ....................................... 60
  6.2.1 Heuristics and Cognitive Biases .................................. 61
  6.2.2 Functional Fixedness ................................................... 63
6.3 The Principle of Autonomy ................................................. 65
  6.3.1 Definition .................................................................. 67
  6.3.2 Jean Piaget and The Rules .......................................... 69
  6.3.3 Autonomy and Functional Fixedness .............................. 73
  6.3.4 Consequences of a Strict Definition .............................. 73
  6.3.5 Heteronomy and the Normative Matrix ......................... 75
  6.3.6 Autonomy and Understanding ...................................... 77
6.4 Arguments against Autonomy ............................................. 78
7 The Human Context ................................................................. 79
  7.1 Cognitive Processes versus Rituals ................................... 79
  7.2 Expertise and Being Human in Processes ............................. 80
  7.3 Morality ......................................................................... 81
    7.3.1 Moralizing ............................................................... 83
    7.3.2 Moral Judgment ......................................................... 84
Part III: Theorizing about Artifacts .......................................... 89
8 The Theorizing Process ............................................................ 91
  8.1 Refocus ......................................................................... 92
  8.2 What Is the Problem? ...................................................... 93
  8.3 Design as Applied Ethics ................................................ 94
  8.4 Conceptions of Good Design ............................................ 94
  8.5 Allowing Autonomy .......................................................... 96
    8.5.1 Aspects That Reduce Autonomy .................................. 96
    8.5.2 Allowing Designers and Developers to be Autonomous .... 101
  8.6 On Theorizing .................................................................. 101
  8.7 Theory of Use Protocol ...................................................... 103
    8.7.1 Design is not Experimental Psychology ....................... 107
    8.7.2 A Hypothesis ............................................................ 108
    8.7.3 Find Meaningful Idea ............................................... 111
    8.7.4 Theorizing ............................................................... 112
    8.7.5 Let the Intended Users React ...................................... 113
    8.7.6 Make Hypotheses ....................................................... 115
  8.8 Risks ............................................................................. 117
    8.8.1 Stress ....................................................................... 117
    8.8.2 Errors ...................................................................... 117
    8.8.3 Biases ..................................................................... 117
  8.9 Related Work .................................................................. 118
8.9.1 Cognitive Work Analysis ............................................. 118
8.9.2 Other Concepts .......................................................... 118
8.10 A Note on Quantitative Data ........................................... 121

9 Examples of Theories of Use .............................................. 123
9.1 Typing on a Touch-Screen Tablet ..................................... 123
  9.1.1 Model ...................................................................... 123
  9.1.2 Qualities .................................................................. 124
  9.1.3 Null Hypothesis ....................................................... 124
  9.1.4 Hypothesis 1 ............................................................ 125
9.2 Smoke Detector .............................................................. 126
  9.2.1 Model ...................................................................... 126
  9.2.2 Qualities .................................................................. 127
  9.2.3 Null Hypothesis ....................................................... 127
  9.2.4 Hypothesis 1 ............................................................ 127
  9.2.5 Hypothesis 2 ............................................................ 128
  9.2.6 Hypothesis 3 ............................................................ 128
9.3 Empirical Points ............................................................. 129

10 Making Decisions Together With Others ............................. 131
10.1 The Ultimate Decision Support, Revisited ......................... 131
10.2 Ethical Decision Making ............................................... 132
  10.2.1 Ethical Competence ............................................... 133
  10.2.2 Related Work ......................................................... 136
10.3 Qualities of a Tool for Ethical Deliberation ....................... 138
  10.3.1 Theory of Use ........................................................ 140
10.4 Autonomy-Oriented Analysis ......................................... 140
  10.4.1 Support Self-Scrutiny ............................................. 141
  10.4.2 Bias Countering ..................................................... 142
  10.4.3 Support Dialogue .................................................... 143
10.5 Support Deliberation ...................................................... 143
  10.5.1 Support Construction of Arguments ......................... 144
10.6 Hypothesis Testing ....................................................... 144
10.7 Iteration 1: EthXpert ..................................................... 145
  10.7.1 Evaluation ............................................................ 149
10.8 Iteration 2: From Analysis to Collaborative Deliberation .... 150
  10.8.1 Democracy ........................................................... 151
  10.8.2 Rhetoric and Discourse ........................................ 155
  10.8.3 Countering Undemocratic Dialogue ....................... 155
  10.8.4 A Mind Set for Democratic Dialogue ....................... 157

Part IV: The End .................................................................. 159

11 Summary ........................................................................ 161
11.1 Introduction to Autonomy .............................................. 161
Part I: Establishing the Question

In this part I account for my view on science and truth, with a special focus on technology and how we interact with it. In subsequent parts, I write about how to enable reflection free from biases, inclinations, and authorities; and how to construct artifacts that do not limit this unnecessarily. Therefore, I offer here a few, perhaps provocative, definitions of concepts that are used throughout the thesis.
2. Introduction

Over the years I have become increasingly concerned with a phenomenon in society, to the extent that I consider it possible that it grows into a threat against humanity as we currently appreciate it. It is not global warming, and it is not overpopulation of the planet, although these are also pressing concerns. Neither is it the risk for nuclear war, nor that of a global pandemic. Not even the ecological consequences of agricultural mono cultures worry me to the same extent.

Such problems present themselves as obvious threats (at least they will eventually), and I therefore assume that the collective initiatives of humanity will eventually start addressing them before all hope is gone. I believe this to be possible, despite humankind’s tendency to act opportunistically. What I’m more worried about is that threat which presents itself as an opportunity; the threat of satisfaction.

How can satisfaction possibly be a threat, the reader may wonder? Especially if you are active within the area of Human-Computer Interaction (HCI), you have probably adopted the idea that satisfaction is one of the core values to strive for in the design of products. The often-used ISO definition of the term is: “Freedom from discomfort, and positive attitudes towards the use of the product” (ISO 9241-11 1998, p. 2).

It is actually not as much satisfaction in itself that I worry about, but what it can do to our thinking. It does not require much argumentation to claim that full satisfaction, as standardized above, implies a state of mind when we do not wish for anything more from the product at hand. We do not hold any negative feelings about it and we are completely free from discomfort. So how can this be something bad?

I do not claim that all satisfaction is necessarily bad; I claim that there is positive and negative satisfaction, and I wish to point out the difference between these. The former, I claim, comes from feelings of accomplishment and self-actualization and the latter corresponds more to contentment and convenience. To get introduced to this line of thinking, consider first a future scenario; a scenario that for every small advance in artificial intelligence and data analysis seems increasingly likely, even though it may still seem a bit like science fiction. It is the idea of The Ultimate Decision Support System.

2.1 The Ultimate Decision Support System

In one possible future, we will no longer make bad decisions. We are past the time of frustration over paying ten coins for something that we could have
gotten for five, we no longer buy products that turn out to be junk, and we
do not marry people that we want to divorce a year later. The evolution of
price and quality comparison services has lead to a situation in which it is
not reasonable for companies to market products that are not among the very
best in some aspect, and we have stopped worrying about the privacy issues in
exhibiting the entirety of our life patterns to dating and entertainment services.
We have never experienced any harm from this, and we are satisfied with what
we are offered because it always turns out to be what we actually wanted.

Different categories of customers, of course, care about different qualities
in services and products, but for a long time nobody has decided anything
without first consulting one of the decision support services. This procedure
has become streamlined to the extent that whenever someone is about to de-
cide something, these services seamlessly inform her what is a suitable option,
based on her—and similarly-minded people’s—previous decisions and prod-
uct reviews. This has come to apply to anything from movies, music, litera-
ture, and technical artifacts, to household produce and even political prefer-
ences. Advances in motivational research have made it easy to design tasks so
that they appear entertaining, even those tasks that used to be experienced as
tedious chores. Without increased effort and with a considerable element of
fun, customers and citizens make seemingly informed decisions.

The peculiar side effect from this servile environment is that people tend
to develop an increasingly consistent taste. After all, they are only exposed
to a selection of services, products and opinions that they likely would ap-
preciate. Reactionary philosophers and psychologists warned that this con-
sistency was reason for worry, as it could create polarized subcultures within
societies. They feared that these subcultures would develop incommensurable
value systems and that their members therefore would end up in conflict with
each other. However, nothing of this sort has happened. In fact, the differ-
ences are not even a topic for discussion, as they are considered as matters of
preference rather than as ideological positions. Different preferences are not
the result of deliberation but of some arbitrary familiarity, in the same way as
your mother’s cooking is likely to satisfy your palate.

This society is not like Orwell’s 1984. There is no oppression of thought
going on. Everyone is free to act and decide how they wish. Big Brother is
benevolent and wishes for you to be satisfied. There is, however—as a con-
sequence of perfect support—a gentle suppression of thought, simply because
people are never forced to deliberate over matters and never have their prefer-
ences questioned. People have learned that their support systems are always
right so why would they make the effort to second-guess them?
2.2 Social Media as an Extension of Capabilities

Consider another example: the popular social media platform Facebook. It is very easy for people to become virtual friends with colleagues, acquaintances and other brief encounters. These people will see some or all of what you share, in their so called “news feed”. Most would agree that this is nice. It gives you the possibility to let others know about your important moments, like your vacations and breakfasts, your outrage at government agencies, and the pictures of cats that you find adorable, as well as to update yourself about what has happened in other people’s lives.

One way to regard this news feed is that it is a good source of information: people share what they think is important and show appreciation for what others have shared. It facilitates dialogue and socializing and we tend to appreciate that.

Another way to regard it is as a tool for surveillance. I do not mean the type that governments are criticized for (e.g., NSA), but the type that neighbors typically engage in. We make judgments about other people based on what we observe. In a neighborhood setting most people have realized that this is going on and act accordingly; we wear clothes when we are outdoors and we greet nicely even when we are in a bad mood; rather effortlessly we appropriate our behaviors in front of those familiar strangers. On Facebook, many have adopted the same self-conscious way of expressing opinions. This is all natural and part of being socialized and is not the problem that I am targeting. However, it creates expectations, and what I see as a potential harm is that the ease with which certain actions are performed may lead people to act beyond their capability to comprehend the effects.

One example of this is the share function that the platform offers. Below status updates, photos, links, and other content there is often a text link that makes it possible to share it with friends. This is very convenient but it does not really communicate what it does, even though it, for an engineering mind, is technically consistent. If the link is under a status update, the status update is shared and you can comment on it. If it is under a photo or a link, the photo or the link is shared and you can comment on it. Easy, right?

The problem occurs when you want to share, for instance, someone’s critical remarks on some content, for instance a link, and hit the share button. You may easily get the impression that you are sharing the critical remarks along with the link but that is not what happens. Only the content is shared. If you do not make a critical remark by yourself, your Facebook “friends” will see you sharing content that you actually oppose. This can become even worse if you add an encouraging comment, intended to support the one who is critical but then instead becomes attributed to the original content. Furthermore, after you have shared your content, you will not necessarily see it in your own news feed so that you can correct your mistake. Your shared content will show up
in other people’s feeds, forming an image of you that you would not be com-
fortable with, without you noticing that this has happened.

This could be lazily framed as a usability problem; that the effectiveness of
Facebook is lacking: you had an intention but the sequence of actions offered
by Facebook led you wrong so that you were not able to accomplish what
you wanted. It is important to realize that this is an evaluative statement. It
determines, after the fact, that there is a problem in the system. I do not claim
that this type of knowledge is unimportant—it really matters—but I claim that
it does not give much input to design; it does not give clues to how it should
have been designed.

Another framing of the problem is that the system does not follow the user’s
mental model. Later in this thesis I will discuss how mental models relate to
the problem that I am addressing. For now, to get back to the discussion about
satisfaction, we can just conclude that it is perfectly possible to be momentar-
ily satisfied by the easy content sharing.

I propose a third way of framing the problem: it is a consequence of allow-
ing a user to act without necessarily understanding what she is doing. This
is a rather controversial statement in many HCI quarters. Do I suggest that
everybody has to understand how a computer operates internally in order to be
allowed to use one?

I will leave this rhetorical question unanswered for now. What I find a
bit disturbing, though, is the impression that people can easily accomplish
things that go beyond their understanding, and that they have little means and
incentives to form an understanding. The current world order expects and
assumes efficiency from working people, on the expense of insight. We are
parts in a machinery that runs ever more efficiently, which is probably good
in some way. We should, however, realize that we are trading in some of our
autonomy in the pursuit. Humans have “always” depended on a system of
people and machines to make life less horrible, but we seem to have become
like lazy emperors that can appoint our subordinates to do the complicated
things for us. We regard computing artifacts as extensions of our cognitive
and motor abilities, and use them to become more efficient and effective in
mundane, as well as sophisticated endeavors. What I fear of losing in this
rationalization process, is our ability to cope with unexpected events, which
requires understanding of underlying mechanisms.
In this chapter I will argue that facts and truths are useful tools to communicate information that has been evaluated, but that they can also become obstacles in understanding the world if they are treated too dogmatically, thus guiding thinking too strongly.

3.1 Induction and Deduction

Within science it is common to contrast two different methods for reasoning: the inductive method and the deductive method.

The inductive method is what empirically oriented researchers do. It refers to a bottom-up approach where we make observations and, based on those, proceed to make general claims about matters. When these claims (hypotheses in scientific language) are tested without being falsified, we tend to become more certain that they are not false and eventually elevate their status to what is commonly known as a theory. Meticulous scientists are, however, aware that theories are not truths, since it is impossible to prove anything general by induction. By merely observing that the sun has risen from the horizon every morning for at least 3000 years,\(^1\) we can still not be absolutely certain that it will rise again tomorrow morning. The logic behind this can be illustrated by an example of modus tollens, an argument form that is commonly used in deduction:

- If the headache pill that I just took works, my headache will be gone in 30 minutes.
- I still have headache after 31 minutes.
- Therefore the headache pill did not work.

This is logically sound. If we, on the other hand, would accept the result that the headache is gone as proof for the pill working (modus ponens), we would commit a logical fallacy called **affirming the consequent** (Popper 2005). The headache could very well have disappeared for other reasons than taking the pill. We thus have no solid proof that the pill really works, we have just failed to prove that it does not.

However neat this reasoning seems, it is not really useful in practice, since what we really want is a pill that works, not one that has no effect. Consequently, induction is about committing the logical fallacy of affirming the

---

\(^1\)I do not want to deter evolution skeptics quite yet.
consequent many times, by repetition reducing the probability that the observed effect is due to something else, like chance. The repetition is done until we have enough observations to dare to claim something. Most of the time, however, this insight is understood only in terms of statistical significance; as a number from a table that suggests a level of certainty that the effect did not occur due to sampling error.

To be fair, also when the pill obviously did not work, it may have been due to an infinite number of reasons, of which the formula of the pill is only one. The empirical world just does not follow the laws of logic very well.

This uncertainty opens up for a spurious type of research, where hypotheses are adjusted to fit the data (Gigerenzer 2004). Given that the ritual of significance testing does not depend on what hypotheses you pose, as long as your hypothesis singles out a pattern from your data set, it is possible to fabricate a publishable, i.e. a statistically significant, result. If you have to reject your original hypothesis about the headache pill but notice that it seems to work better for a certain group, for instance pregnant women, you may be tempted to adjust your hypothesis, despite lack of theoretical justification, in order to get statistical significance. The result is a seemingly publishable, yet highly questionable pseudo-scientific paper.

The deductive method, inversely, is a top-down approach where some truths are asserted and claims are logical consequences from these. It is what computers, mathematicians and many philosophers do. In the case of the sun, we must be able to assert that the earth has a constant revolution around its axis and follows its path around the sun, and that the sun will always be there, to claim that the sun will rise in the horizon in the morning. Given that this is a statement about the empirical world, it is not really appropriate as an example; it should not be possible to refute the premises. The syllogism should only be invalid if the conclusion does not follow from the premises.

But we could, for instance, define morning as when the sun rises over the horizon. The modus tollens is a valid syllogism:

- If the sun rises, it is morning.
- The sun does not rise.
- It is not morning.

And so is modus ponens:

- If the sun rises, it is morning.
- The sun rises.
- It is morning.

Essentially, induction and deduction build on the same system of logic. The difference is that in deduction we assert the premises and draw logical conclusions, while in induction we test a premise by observing the conclusion.
In practice, both methods are necessary. If we would define every particular phenomenon in order to allow deductive statements, we would not be able to make any conclusions about phenomena that we have not yet defined. Here enters scientists’ interest in finding inductively developing theories that are possible to generalize over a range of phenomena. Theories that explain many types of phenomena reduce complexity, as we can then assert less to deduce more.

3.2 Abduction

Abductive reasoning, i.e. inference to the best explanation, is the fundament for making predictions in real-life settings, including research processes (Douven 2011). The term was coined by Charles Sanders Peirce in the middle of the 20th century and is meant to denote the logic of making some observation and then engaging our imagination and experience to explain this for ourselves. Assume, for instance, that you come home and find the fridge door open. You have no recollection of leaving it open so you start to think about what explanations there are for this unpleasant surprise. You will probably try to close it to see if there is something preventing it from closing, you will open it again to determine if it might have opened by itself. If you cannot find a physical explanation, you will proceed to process other likely scenarios. Without any additional proof, there will probably still be some explanation that seems more plausible than others. If you have a daughter who has left the fridge open before, that is probably your preferred hypothesis. If you are living by yourself, you will probably blame yourself, after verifying that nothing is missing from the fridge and the rest of the apartment. Other explanations exist and are equally supported by the thin evidence but they do not seem very likely. This is abductive reasoning at work.

In this thesis, I assume that this is how the human mind works. All psychological concepts that I will present, and the procedures that I will suggest should be interpreted with this in mind. Abductive reasoning allows us to be creative as well as conservative. Tolerant as well as intolerant. It all comes down to what stimuli we get; what we determine to be the most plausible explanation for an observation.

This type of reasoning, however, can also lead wrong sometimes. Studies by, for instance, Amos Tversky and Daniel Kahneman (1974; 1983) have shown that it is easy to construct cases where this ability to imagine violates the laws of logic. We should, thus, keep in mind that what is plausible is not necessarily true, but if we are careful, it helps us to discover things that would be impossible with only deductive and inductive methods.
3.3 Logic

Logic is that which forms beliefs into coherent systems. When I studied boolean algebra, I was taught that true is the negation of false and vice versa. This was postulated as axioms. It seemed rather trivial at first, as I was judging this piece of truth based on my common sense understanding that something cannot be both true and false at the same time. It is something that every child learns in school and for many years I had used this piece of knowledge in programming, without ever reflecting over it. I saw logic as objective and just something to be discovered. Unconsciously it had permeated my thinking. This presumption was effectively deconstructed when I read about the Stoics’ struggles to formulate the logical axioms that we now take for granted. Some systems of logic, that existed prior to this work seem naïve today.

One documented example of this is Philon’s view on hypothetical propositions as based on the current state. This implies that a proposition can be false only if the premise is observably false (Marc-Wogau 1970). The proposition “if it is day, it is night” is, consequently, only false during daytime. This line of thinking, although unfamiliar to us, may still have been functional. During night time, we cannot know for sure what the conditions will be during the day so we better regard statements about it as true. It seems based in the idea that true is that which we cannot reject empirically. We can no longer liberate our rationality from the logic that we know. It forms the foundation of our thinking and we cannot “unknow” it.

Humans can learn to deal fairly well with some logical claims but not all. I will borrow a few examples from Johnson-Laird (1983) to make a point about the logic of humans. Consider these two statements:

Some of the artists are barkeepers.
All of the barkeepers are chemists.

It is pretty easy to make a valid syllogism: Some of the artists are chemists. Also the converse formulation makes immediate sense: Some of the chemists are artists.

On the other hand, there exists syllogisms that are more difficult to make. Consider:

All of the bankers are athletes.
None of the councilors are bankers.

What conclusions can we draw? This is more difficult than the previous example. Johnson-Laird’s theory is that the difficulty of a logical problem is dependent on the number of mental models\(^2\) that is required to exhaust all possibilities. The more mental models we need to entertain simultaneously, the

\(^2\)Johnson-Laird uses the term mental model a bit differently from how I use it in this thesis. He focuses on the mental mechanisms that constitute understanding—what it is and how it
more difficult the problem becomes. He refers to experimental data to support this claim. However, accepting the same data, I would frame the conclusion a bit differently: as the introduction of “maybes” on top of the unknowns. I call those syllogisms maybes that are abductively implied by the statements but not exhaustively determined: some councilors may be athletes, some athletes may be neither bankers nor councilors, some councilors may be athletes.

These statements make abductive sense but are not determined logically. Amidst all these maybes, we have a hard time to realize that all we can say for sure is what the first and second premise give unambiguously: some athletes are not councilors. Introduction of maybes makes it difficult for us to maintain strict logic, since “maybe” is not part of deductive logic, while it is part of our reasoning. For this type of problems, Philon’s logic seems more familiar: all those maybes could very well be true since we cannot know for sure that they are not.

3.4 Science and the Truth

I think that we shall have to get accustomed to the idea that we must not look upon science as a ‘body of knowledge’, but rather as a system of hypotheses; that is to say, as a system of guesses or anticipations which in principle cannot be justified, but with which we work as long as they stand up to tests, and of which we are never justified in saying that we know that they are ‘true’ or ‘more or less certain’ or even ‘probable’. (Popper 2005)

Not even science\(^3\) is immune to the logic of abduction. The espoused aim is to seek truths but this endeavor is heavily constrained by the truths that have already been established. Scientists may pay lip service to strong arguments suggesting that we—as fallible humans with limited cognitive and perceptive abilities—are incapable of observing anything with absolute confidence, while—as socialized, mortal humans—still seeking refuge in estab-

\(^3\)What I mean with the term science is sometimes referred to as wissenschaft in anglo-saxon literature. I still rather use the simpler of the words to refer to the endeavor of generating new theories. I know that the term is often limited to include only certain material sciences like physics, biology and chemistry but I cannot appreciate such a distinction as I believe that it hampers the possibilities to generate useful ideas. If scientific status is guaranteed within certain disciplines, it leads to dogmatism within these, which also guarantees that unscientific practices will flourish, and if it is unattainable in other disciplines, like for instance sociology, the tradition does not contribute to self-criticism. If, on the other hand, all scientists, of all disciplines, were expected to justify why their efforts are science, i.e. contributions to the global “body of knowledge”, we would probably see more appreciation for, as well as actual scrutiny of, other disciplines’ research problems.
lished “facts”. Very few would stand the solitude and insignificance that fol-

ows from admitting that everything is open for interpretation, even if this
would possibly lead to a better chance of truly understanding the world.

I admit to having a tendency toward relativism but I reject neither knowl-

edge nor human values and I am deeply skeptical toward phenomena that
I cannot find rational explanations for. I would not claim that I know the
truest truth, at the same time as I do not agree with the proverb of Paul Fey-
erabend (1993) that the most constructive attitude for science to progress is
“anything goes”. For that, I put too much value on good theorizing and argu-
ments that I consider to be logical. To me, science is theorizing. But theory
should be useful, not just a comfortable narrative for structuring past events.
A good theory has tool value, and perhaps it should be demanded from it that
it makes it possible to formulate predictions.

Theories are very useful for articulating, in an understandable way, what is
happening in our worlds. However, theory is not only explanatory; it is also
shaping our cognition; a strong theory reinforces what is important.

If a theory would have only explanatory and no predictive power, it would
not help us make more sense of what is happening in the present, while still
limiting our scope of imagination. We would only be able to narrate what has
already happened in a specified setting, while we ideally want theories that
help us make good decisions and avoid bad ones. However, I do not claim
that the predictive power needs to be expressed in quantifiable terms. Theo-
ries about how to design an artifact are sufficient if they give the foundation
for—predict—good design. In Part III, I define theory in a way that enables
systematic theorizing about human-computer interaction.

3.4.1 The Truth

When Galileo proposed that he had proof for the Copernican world-view, the
catholic church forbade him to disseminate these findings. What is less known
is that a committee of experts determined Galileo’s proofs to be unscientific
according to the scientific standards of the day. The instrument that Galileo
was using to make his claim, the telescope, was unproven, and it could not be
taken for granted that it gave a true representation of the sky. After all, what
could be observed with the eye differed from what could be observed through
the telescope. Not only was Galileo’s claim unscientific, it could potentially
do harm to society by false beliefs being spread (Feyerabend 1993).

All that we regard to be true is defined as true by humans. This may seem
like semantic nitpicking, but what I try to suggest is that truths essentially are
true by agreement. $T = \neg F$ is defined to be a true equality. “I am human” is
equally defined, linguistically. “The sky is blue” is also true by agreement, and
tacitly we know that we are then talking about a clear sky during daytime. The
same applies to expressions like “thou shalt not kill”, “we stick to the surface
of the planet because of gravity” and so on. These are useful conventions and we should be grateful to have them, but they are nevertheless only true because we have agreed on it.

You may think that this type of relativization of objective truths is extreme. I can imagine arguments like “we know that truth is not objective, we just use the word because it conveys what we mean”, in a similar way that Karl Popper distinguishes between truth and certainty. My answer is that words like “truth” and “fact” do not just convey a conviction; they contribute to determining it. When we use these words, we start believing that what we state is objectively true. It becomes a rigid part of our understanding of the world, of our ontology.

Indeed, truths are of great use. They really do not need to be objective, in the sense of existing outside of human thinking, they only need to appear purposeful. Truths are social phenomena: constructs. An hermit would not necessarily need to define them as such, since the truth attribute fills a purpose only in communication. For the hermit, things may very well just be in a certain way; considered as matters of fact, whereas social encounters necessitates a somewhat shared understanding of the world; the common ground that enables communication. Even if the hermit would try, she would not have any chance to verify a truth claim as there is nobody else around to confirm, or argue against, it. It would thus be pointless to try to establish as a truth that all red berries are poisonous. However, this does not prevent her from acting according to beliefs that she has established for herself; it is just the attempt to universalize one’s beliefs that is futile.

Finally, to attach the “objective” property to truths is only more purposeful when we feel the urge to convince others that they are wrong. It is thus only in persuasion that objectivity matters. This is, for apparent reasons, a very common strategy in western science and philosophy (Slife and Melling 2012).

### 3.4.2 Objectivity

Objectivity is institutionalized subjectivity. Objectivity is achieved by theories and models that reduce the influence of individual affects, biases, and opinions. It has many advantages in many scientific contexts; it reduces the risk of unconsciously selecting only data that support a proposed idea, it allows researchers to replicate experiments, it reduces the cognitive effort of justifying a research approach, and it limits the risk of charlatans posing as researchers.

But should we always strive for objectivity, which in essence is objectification of subjects? Is it not the case that we are putting all of the proverbial eggs into one basket when we bypass individual judgment, in favor of models that by design are not perfect, although often very useful? What happens when one of those anomalies, that seem to be more common than we want, strikes a system that is built to disregard it (Taleb 2010)?
Still, without objectivity as an ideal, everything would be case based, and all answers would depend on who poses the question. It would not be possible to refute claims, which is crucial for the development of science: without formulating general statements—if everything is dependent on the context—nothing can be assumed.

This of course reflects a humble attitude to life, but it also invites unreflective, rhetoric statements about the matter of things in particular instances, which in the long run accumulates to a pile of particulars. There is great value in rich descriptions of particular situations, but it would also be valuable if such research results could be communicated in a form that contribute to the general body of knowledge within a field. Within Human-Computer Interaction, this implies striving toward finding the essences in use situations, the features that can be generalized to hold valid in different settings.

3.4.3 Values in Science

Despite many efforts to prove otherwise, values are an intrinsic part of science (Rudner 1953; Bergström 1994). Every aspect of the scientific enterprise is affected by human decision making, based in subjective assessment of what is important to consider and how to do this best.

Our values are determined by our ontologies—what we perceive to exist in the world and how this functions—which in turn makes us unable to appreciate the value of concepts that are not part of this. For instance, if a deity is not part of a person’s ontology, that person is incapable of experiencing the world in the same way as a religious person. We need to become sensitized in order to read information in what otherwise is just meaningless symbols; data. Analogously, we cannot avoid being influenced by our conceptions when we are interpreting phenomena. They make up the matrix through which we perceive things.

The logic that I have obsessed about is a property of my subjective world. I do not rule out that there may be other, equally arbitrary yet ontologically consistent, logics; self-contained systems in a sense. The problem is that I will not be able to appreciate these from my point of view and I will most likely describe them as dogmatic in the context of my own ontology. Religion is one such partly overlapping yet alternative ontology, which includes the existence of deities. Seen from the outside, ontology constitutes a set of dogmas. Seen from the inside, it is a natural existence of phenomena. By advocating a liberal attitude toward conflicting ontologies, a more humble communication could follow.

---

4I realize that I use the concept ontology very liberally. I believe that using it to understand to differences in values is a way to make concrete practical use of a disputed philosophical term.
4. Technology

The extension of any one sense alters the way we think and act—the way we perceive the world. When these ratios change, men change. (McLuhan and Fiore 2001)

The making and use of artifacts is very much a part of all human endeavor, and so what is to be made and used is primarily a moral question. Technology is a subdivision of ethics. (Feibleman 1967, p. 333)

Do humans control the way technology develops or does technology evolve independent of human will?

The latter idea is called technology determinism. Within it, two main categories can be distinguished: those who think it is positive for us, and those who fear that it is detrimental for the survival of the humankind. Those who are positive, anticipate a “transhuman” world, in which our minds are gradually merged with technology to finally end up in a complete symbiosis. Others see a future in which technology is taking control over our lives, to eventually replace us, or a future in which we use technology to extinguish ourselves.

The historian of technology David Nye does not believe in either kind of technology determinism (Nye 2006) but he readily assigns others these positions. One of these others is Marshall McLuhan, the controversial media philosopher. Technology determinism is, however, not an easily classifiable position and others strongly object the accusation that McLuhan would be a representative of it. Nevertheless, he has written in a way that could be interpreted as a belief that technology has its own agency:

Print technology created the public. Electric technology created the mass. The public consists of separate individuals walking around with separate, fixed points of view. The new technology demands that we abandon the luxury of this posture, this fragmentary outlook. (McLuhan and Fiore 2001, p. 69)

Rather than attributing agency to technology, I understand McLuhan to point out that technology is an emergent effect of mindless “progress”. Technology, as a faceless sheep dog, makes it apparent that humanity as a whole lacks direction despite plenty of individual ingenuity. This interpretation is supported by other claims. “[T]here is absolutely no inevitability as long as there is a willingness to contemplate what is happening” (McLuhan and Fiore 2001, p. 24).
The views on technology development that I am advocating in this thesis may appear a bit negative so I want to stress that there is a difference between designing for autonomy and the Luddite view that technology is predominantly dangerous. Although I am skeptical toward automating decision, I do not contend that technology has tool value. Almost 30 years ago, Hubert and Stuart Dreyfus voiced the very same concern:

> We are in danger of becoming a society that confuses computer-type rationality with true expertise. If we fail to put logic machines in their proper place, as aids to human beings with expert intuition, then we shall end up servants supplying data to our competent machines. Should calculative rationality triumph, no one will notice that something is missing, but now, while we still know what expert judgment is, let us use that expert judgment to preserve it. (H. Dreyfus and S. Dreyfus 1986, p. 90)

Since then, technological development does not seem to have taken a different route. Despite widespread awareness of the potential risks in substituting domain experts with technology, this is what is happening every day. One of the latest professions to be challenged is that of the physician. Research is conducted about automating aspects of clinical decision making. Many physicians have adopted a practice of consulting online support systems when making diagnoses (Franko and Tirrell 2012). Such initiatives and practices have the laudable aims of helping more people and reducing misdiagnoses, but the development should be accompanied by attention to how this type of support affects the professional judgment of practitioners.

> At the core of these values is the notion that technology should serve human intelligence and imagination (rather than the opposite), and that people’s experiences with technology should be structured in accordance with their abilities of perception, cognition, and movement. (Cooper, Reimann, and Cronin 2007, p. 150)

Technology should not make the human efficient. It should be efficient for the human operator to use. The difference may seem trivial but the implications are great. If technology is designed based on the former ideal, in which a human and machinery behave like a unified system, technology may end up controlling the human as much, or more, than vice versa, and it can still be considered to be successfully designed. If the latter is the ideal, the evaluation of technology must include the human’s point of view, in terms of usability, controllability and appropriateness, and cannot be based only on human-machine-system output.

### 4.1 Artifacts

An artifact is an object that is intentionally produced for some purpose, by a human (Hilpinen 2011). In other words, artifacts are materializations of hu-
man decisions. Humans have created some thing, to overcome some obstacle, to get to something.

Joining Hilpinen (2011), I will in this thesis use artifact as a term for everything that is intentionally made by humans. This extends the view that artifacts are material, represented by, e.g., Feibleman (1967), to include also immaterial objects like software. I prefer this term instead of technology, software, device, gadget, etc., because it does not bear connotations of modernity or utility. Everything that has an author, to put it a bit poetically, is an artifact, which includes the above things. I should, however, be careful. By grouping qualitatively different things into the same category, there is a risk that I end up revealing nothing at all; that the category becomes useless for analysis. There is also a risk that my claim will be falsified by using an instance of artifacts that does not fit the pattern that I assume. These are real risks, but I consider the value of pointing out that something is made by humans, and thus also malleable for humans, worth the risk of eventually having to narrow down the definition.

I will extend Hilpinen’s notion a bit by pointing out a peculiar characteristic of artifacts. Some of the things we have in our lives are obvious materializations of rules, to control ourselves and other people: speed bumps (Verbeek 2006), fences and locks come easily to mind. Another, less obvious example is the design of toilets. The philosopher Slavoj Žižek has humorously suggested that the design of toilets reflects cultural attitudes. One of the examples to support this claim is the fact that German toilets traditionally were designed with a shelf onto which the excrement will land, rather than the strategically located water surface that is common in most western societies. Žižek used this design decision as evidence that Germans are culturally analytic and, as such, interested in continuously examining their faeces for possible anomalies. I do not wish to delve into that matter, but just suggest that another phenomenon can be observed around this particular design. Even though these toilets are now, to an increasing extent, replaced with the less explicit version that lets the odorous outcome submerge in water, a cultural practice has survived the transition, namely the inappropriateness of men standing while urinating. This norm is somewhat emerging also in Sweden but in certain parts of Germany it is, according to anecdotal evidence, a very boorish practice, which is obvious if you consider the effect of a jet of water on a flat surface. What is curious about this, is that the impression of the urinating style has survived even though the practical effect has changed.

What I claim is that artifacts to varying extent function as rules; rules which then can live on as norms even when they no longer have the properties that enforced a certain behavior. One way to put it is that artifacts can be moralizing; that they shape our behaviors (Verbeek 2006); that they form rules for

---

1This is, again according to anecdotes, not the case in the Netherlands or Austria, where toilets are designed in the same fashion.
us to follow. This may seem like an unnecessary semantic exercise but I think there is value in actually defining, rather than having implicit, what an artifact is.

When it comes to rules, they are generally something that determines the proper way to accomplish something. They constrain our choices at the same time as they supply something reminiscent of a guarantee of predictability. This is useful in societies, as we can have a good guess about how other people will behave.

The above observations are not enough to claim that all artifacts are necessarily reified rules but we can safely assert that both categories are the result of some decision-making process, and that there are some shared properties. Now, does it matter whether a norm is enforced by an artifact or by a social rule?

A problem of moralizing artifacts is that they ruthlessly intrude on human autonomy. The more we are directed by entities in our environment, the less free we are to follow our own reasoning when we choose how to act. Verbeek tries to counter such arguments by claiming that we are anyway controlled by laws in society: “Our legal constitution implies a major limitation of freedom, after all, but this does not make it a threat to our dignity” (Verbeek 2006, p. 370).

However, in societies, laws (read rules) are enforced by humans, and violations are handled by a deliberating body: the justice system. This justice system is obliged and capable to make an informed decision about how severe your violation was. If you ran across a lawn where walking was not allowed in order to save a drowning person (see Figure 4.1), that is a negligible violation. If this, on the other hand, was a rule that was enforced by an artifact, e.g., a fence or moat, you would not be able to violate it and you would not be in the position to choose your actions. When it comes to computerized artifacts, the norm seems to be to make rules that are impossible to violate. The law scholar Lawrence Lessig (Lessig 2009) argues that software should be seen as rules that are designed for that which has been anticipated by the designers. This is in line with Zabban’s reasoning about computer games as “frozen discourses” in the sense that these games legislate how to play them (Zabban 2011).

Nevertheless, if we assume that artifacts do exert some control over our behavior, we should be aware of this and pursue design that is moralizing intentionally rather than as an unanticipated side effect. In this I agree with Verbeek. For Verbeek it is important to democratize the design process by inviting relevant stakeholders to contribute to the conceptualization of an artifact. I will later argue for a way to accomplish this by focusing on designer and user autonomy.

Some artifacts extend our ability to imagine what is desirable. Works of art are suggestions that there might be something else to desire. Katherine Leduc suggests to regard artness as an affordance of an artifact, in Norman’s earlier sense; that it is a property that can be perceived or not (Leduc 2013).
rather perceive art as exhibiting affordances in the world, with the purpose of revealing and affecting what it means to be human. In that light, I pretentiously try to produce arguments for designing artifacts that help people to imagine a world in which humans are more important than systems.

By posing rules, artifacts shape cognition. They affect our possibilities to reason and this is especially true for cognitive artifacts, like computerized tools (Woods 1998). This means that there are moral implications in artifact design, a fact that is probably not discussed enough among designers, engineers and developers. Artifacts are affecting the system of people and we should be aware of how.

Another noteworthy point is that an artifact is a product of a creator’s ability to imagine. “The existence and some of the properties of an artifact depend on an author’s intention to make an object of certain kind.” (Hilpinen 2011). The cognitive and computer scientist Gerald Sussman says that computing is done wrong when we are trying to design explicitly for events that can be foreseen by designers. Systems should instead be organized so that the consequences of design decisions are not expensive to change if these decisions are proven wrong (Sussman 2011). This is an ideal to keep in mind and I will below propose a way to approach this in a systematic way.

4.2 Tools

When I look around at instances of good system design—systems that I think have had profound influence upon the users, I find that what seems more im-

Figure 4.1: If rules were not possible to violate, it would not be possible to save the drowning person.
A special type of artifacts are tools. When trying to find a dictionary definition of the term, these typically describe and classify concrete, physical working tools: knives, hammers, screwdrivers, power drills, chain saws, etc. (tool 2014a; verktyg 2014; tool 2014b). The encyclopedia Britannica states that a tool is “an instrument for making material changes on other objects, as by cutting, shearing, striking, rubbing, grinding, squeezing, measuring, or other process.” (tool 2014a). This kind of definitions seem a bit narrow, as I suspect that most people today include in the term a wider scope of devices, including computers, cameras, algorithms, computer software, etc.; things that we use more or less frequently in our daily lives. To understand how we have ended up at our current conception, which is not reflected well in dictionaries, I will later try to crystallize an evolution in the field of computers. But first I would like to supplement the above definition by presenting some ideas about what a tool is.

In a short essay referred to as a rant, Bret Victor offers an elegant-sounding definition, general enough to cover a vast range of things, at the same time as it suggests a horizon for what is included:

A tool addresses human needs by amplifying human capabilities. (Victor 2011)

The problem is, however, where to put the horizon. Is, e.g., food a tool? It amplifies our capabilities by supplying nutrition, which at the same time is a proxy for the arbitrary human need survival. Food could also fulfill the need for satisfaction by being tasty and beautiful. Even though the mentioned amplification of human capabilities would be possible to delimit, I am not convinced that we can ever create a satisfactory demarcation of what a human need is. Given an appropriate context, every desire of a human being can be interpreted as one, especially if we include arguably important social and cognitive aspects. Today, many people, e.g., feel a “need” to be able to get in touch with people at any given occasion and, consequently, feel legitimate stress if their phones run out of battery.

A less ambiguous definition would be that a tool is just an amplification—or in Douglas Engelbart’s words: an augmentation—of human capabilities for whatever purpose we wish to use it (Engelbart and English 1968). The perhaps unsatisfactory consequence of this definition is that everything that we can imagine, that affects our physical or reasoning abilities, then could be regarded as tools. Food could definitely be a tool, a thought could be a tool, air could be a tool, a mountain could be a tool, a planet could be a tool, etc. A minimalist definition, suggesting that everything is a potential tool, is in my opinion not necessarily a problem, as tool value is determined in a situation rather than a priori. On the other hand, it can possibly render the term useless
for analysis. This depends on what aspects of the category one is interested in and I think it is an interesting avenue to explore. So I conclude this far that a tool is one of many possible interpretations of an object, that implies that the object is usable—in the word’s most general meaning—to achieve something.

At the same time we might prefer to take the Wittgensteinian stance (as interpreted by Hedman (2004)) that we cannot properly delimit the scope of things that people consider to be part of one and the same category. In the world of Wittgenstein, the word “tool” describes a family of objects, which cannot be uniformly defined but nevertheless have some, but not necessarily all, characteristics in common with other members of the same family. This would render a wide definition of the word, but perhaps still not as wide as the previous suggestion. Some tools would in such a scheme be the classical artifacts that we use to directly manipulate other objects, and perhaps this would be the shared property of these. Their “children” may be tools to enable the use of other tools and perhaps their “cousins” are tools that are used without directly affecting any other objects, nevertheless with some property shared with some other tool. A map could be an example of the latter. I would argue that the idea of family resemblance is useful for describing how the classification of artifacts has evolved, but it does not present an obvious opportunity to categorize artifacts in a specific context, as it is dependent on finding the linkages between different objects, rather than making a judgment by just observing how objects are being interacted with. This means that we would have to search for some commonness or identifying quality.

For a current-day definition, in case such is necessary to state at all, I suggest that one thing that is in common for all tools is that they are seen as tools when they are handled by a user with an intention to act, i.e., to perform some arbitrary procedure. A computer can be just something that people fiddle with, without any purpose whatsoever until someone uses it to, e.g., make calculations or play a game for relaxation; a map can be just a decorative object in the hallway of an office building until someone decides to use it purposefully; and this is probably the insight I want to arrive at:

A tool is something that is used purposefully to perform a procedure.

This definition differs from that offered by Victor in that it does not require identification of needs. In the classic case of the hammer, it seems a bit constructed to imagine a need to drive a nail through a plank rather than regarding that as a procedure that is performed for some higher-order purpose.

Neither is it normative in the sense that a tool has to amplify human capabilities, nor does it suggest that a tool is necessarily always a tool: an object is a tool when it is used as a tool, to put it a bit tautologically. Here we are touching on one important aspect of what I advocate: if we refrain from classifying objects categorically, we are more free to interpret them as we find useful, when we want to. We are, also semantically, allowed to appropriate.
A Utopian Project

Appropriation was important in the UTOPIA project: a collaboration, conducted between 1981 and 1985 (Ehn 1988), between the trade union for graphic workers and a group of computer and social scientists. It became the most influential example of participatory design, an approach that is famous as a “model for participative and democratic technological research and design” (Ehn 1988, p. 342). It was initiated based on concerns that the introduction of computers into the domain of graphic work would reduce the degrees of freedom of the graphic workers, with the consequences that the quality of work would decrease, and that work could be experienced as less meaningful.

In the Utopia project, computers were explicitly regarded as tools, with the sole purpose of enhancing or amplifying the skills of their users. This was based on a design philosophy that regarded craftsmanship as the primary condition for expertise. If computers were not regarded as tools, they could not become ready-to-hand for the user, and could therefore risk reducing her professional skills. The explicit perspective on computerized artifacts as tools was not dominant before the UTOPIA project, and is not dominant today.

In many respects the requirement specifications looked like traditional ones. It was a forty-eight page technical document filled with musts and ought-to’s, and questions to be answered by the suppliers. There was a description of requirements for work procedures, display screens, human-computer interaction, and operations for page make-up and image processing equipment.

What was new was that these technical requirements were derived from the principle that the equipment should serve as tools for skilled work and for production of good use quality products. The requirement specifications were also different in the sense that they also specified requirements for work organization, work environment and training. Technology bringing about poorer working conditions than those already existing was not accepted. (Ehn 1988, p. 339, original emphasis)

As a final remark, we can learn by observing a group of expert tool users, namely musicians, that what is desirable in a tool is not that it always produces a good result, but that it is empowers a skilled user to achieve what she intends.

What we are affirming is that the performer should never be the instrument’s slave and that a good instrument should not impose its music to its player. A good instrument should not be able to produce only good music! (What is good music anyway?) A good instrument should also be able to produce ‘terribly bad’ music [...] (Jorda 2004)
4.3 Designers

It should not be very controversial to claim that everyone is a designer. This conception has been popularized by the well-known polymath Herbert Simon, who offered an often-used definition in his *The Sciences of the Artificial*:

> Everyone designs who devises courses of action aimed at changing existing situations into preferred ones. The intellectual activity that produces material artifacts is no different fundamentally from the one that prescribes remedies for a sick patient or the one that devises a new sales plan for a company or a social welfare policy for a state. (Simon 1996, p. 111)

This view is furthered a bit by Gerhard Fischer (Fischer 2002) who asserts that most people want to be designers of their everyday lives. They want to be in control.

What if it is so, that when we can no longer affect—control—our everyday matters—like in a job in which we do not get to make any actual decisions; or in a political sphere that is occupied by parties that seem insensitive to what people want; or in the market place, where the ethical decisions of an individual do not seem to affect the decisions of multi-national corporations—we become more obsessed with controlling what we can? To continue the speculation a bit, this could, for instance, explain the increasing interest in artisanal expertise. It is (was) fashionable to bake and eat sour-dough bread, it is (was) fashionable to brew and drink craft beer and it is (was) fashionable to do handicraft. These are tangible pursuits, involving the prospect of distinguishable skill improvement. In the technology domain, some people buy Raspberry Pi’s and circuit boards, and wire them up to sensors and actuators to build their own unique technical gadgets. They call themselves “makers” and display a strong desire to appropriate technology for their own purposes. It seems to be less important that the resulting artifacts have utility and more important that they are in fact crafted; designed by the user.

This would suggest that my fears in the introduction probably are unfounded; that people’s wish to be autonomous finds an outlet, despite the lure of convenience that is offered.

There is, however, a great difference between designing for yourself and designing for someone else. In the first case, you are constantly guided by your own vision of the finished artifact, which is further stimulated by the outcome of your continuous efforts. While you are designing an artifact, you are constantly assessing it and you can allow yourself to change your mind if you notice something interesting to investigate further, or if you realize that your initial thinking was flawed. Designing for others means that you need to understand how they think, and how they want to use an artifact. This creates an interesting obstacle, since you cannot trust your own assessment fully. It is a different activity but is by some programmers seem to confuse it with the first type. What I propose below is a way to address this obstacle, without
exerting the type of outside control that reduces the intrinsic quality of being in control.

When I below write about designers, I do not limit it to mean the ones who polish the aesthetics of an artifact. Like Simon, I use the term for everyone who makes decisions about how to shape the future. However, in some chapters—ironically those in which I think this view is especially important—I am not taking the risk of being obscure so I add “and developers” after designer. Also, bear in mind that also users are designers. They may not design the tool that they are using, but if they see a need for it, and are allowed to, they will appropriate it for their purposes.

Another way to regard designers is as craftsmen. In The Question Concerning Technology, Martin Heidegger (1977) used the concept of poiēsis to describe the refinement of material into something that had greater aesthetic and moral value. A carpenter that is taking a piece of wood and brings forth its intrinsic qualities—for instance its color, growth rings and fiber structure—in a piece of carefully and deliberately crafted furniture is engaging in this. When it came to modern technology, Heidegger did not see any poiēsis going on. He considered it as tasteless challenging and exploitation of nature.

But perhaps it is still possible to talk about poiēsis when it comes to designers of technology; as crafting the digital material into a meaningful, tasteful representation of its essence? Many computer scientists do have aesthetic ideals when it comes to programming code, and usually what is considered beautiful is that which is an honest, yet refined, representation of the algorithmic nature of the digital. User interfaces are commonly frowned upon, as these, by introducing patched and idiosyncratic commands to create visual representations that do not reflect the beauty of the algorithms beneath, are disturbing the ideal. Existing frameworks for creating user interfaces are not made for bringing forth the digital material, but rather for putting on a layer to hide it from the user. With the advent of embodied interaction and the Internet of Things, we can, however, come closer to such an ideal, in which the digital can map beautifully to physical representations. The maker movement is indicative that there may be something intrinsically human in that.
5. Human-Computer Interaction

The HCI community should not restrict its efforts to user interfaces or the computer part of HCI confining itself to focus solely on some technical issues in the context of a world defined by others. (Fischer 2002)

This chapter describes my view on the current state of HCI. In summary, I claim that a majority of researchers in the field are too concerned with variations of how to construct better interaction design and thus forgetting to ask themselves why this is done. This is perhaps a heritage from the field’s background in computer science engineering, but that alone cannot explain it. We need to add psychologists’ interest in observable expressions of human behavior to the explanation.

One typical symptom of this misalignment is the quest for methods to study users. The reasoning goes that, in order to make computerized systems usable for intended users, it is necessary to study how the latter behave. Consequently, for instance ethnography is often considered as a method to inform designers about users and not really interesting as a result in itself (Dourish 2006)—at least not interesting enough to motivate the costs associated with producing ethnographies. I would argue that it is interesting to know how people behave with computerized artifacts, even if that would not inform design directly. Based on an understanding of the roles that computerized artifacts fill in people’s lives, we can draw conclusions about whether we are heading in the right direction.

Another symptom of the misalignment is the obsession with evaluation. Evaluation will give a designer a posteriori information about how well a system works. Surely this information can be used to improve a system, or at least to find out what does not work, but it is essentially an engineering practice that is focusing on treating symptoms rather than causes. I claim that an unreflected adoption of evaluation as a best practice undermines appreciation for a priori expertise. A priori expertise must start from understanding why something should be done in a certain way. What constitutes the why is the purpose; neither the observable nor the espoused one, but the one that is intrinsic for humans.

Let me get back to this confrontational idea later and instead proceed with a brief, selective and incomplete survey of the field of HCI.
5.1 Birth and Focus of HCI

There has been interest in Human-Computer Interaction (HCI) for almost 70 years, if we count as the starting point when Vannevar Bush presented his ideas for a device to augment human thinking (Bush 1996). This is an arbitrarily chosen time of birth for the subject; there are other candidates, like when Ivan Sutherland presented the first graphical user interface in 1963 (Sutherland 2003), or the advent of cognitive engineering in the late 1970’s (Carroll 2013). The exact time is not really important. What matters is that the field has grown and changed focus over the years and many would agree with me that the concept of usability has been at the core of the field during its blooming. Let us therefore start by examining this idea.

One long-lasting definition of usability is the one introduced by the International Organization for Standardization (ISO) in 1998. This was the result of many years of deliberation and negotiation, and can thus be regarded as the smallest common denominator for the state-of-the-art at the time. After its introduction, it has functioned as an argument and a mantra for usability-oriented research. According to the definition, usability is:

The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO 9241-11 1998).

Pay special attention to the use of the word “specified” and the three core values: effectiveness, efficiency, and satisfaction.

There are other slightly different, yet in essence similar, definitions highlighting, e.g., learning, memorability, and flexibility (Welie, Veer, and Eliëns 1999), and variations that are stressing, e.g., accessibility (Shneiderman 2000). What should be noted is that these definitions are all goal oriented, in the sense that users are perceived as task performers, and the role of a computerized artifact is to support them in finishing tasks that lead to goals with minimal cognitive effort. However efficient this may seem, it may not map very well to how humans really work. Lately, partly as a critique toward this rather instrumental, tayloristic view of humans, usability has been eclipsed by a greater focus on the overall experience that a user has of a system, tool or product. This is also reflected in subsequent versions of the ISO definitions.

If I have given the impression that HCI is all about usability and similar software centered concepts, I will now try to correct this. Today, there is a large acceptance of what can be published within the field, which is both good and bad. It is good, since it proves that the community is evolving and open to new ideas and perspectives, but it is bad for the theoretical identity and mutual sharing of expertise within the community. It can also lead us younger researchers astray, as we are flooded with epistemologically different concepts before we get the chance to find our own stance. Then again, a strongly rooted identity comes with both positive and negative effects. It may
contribute to a clear perspective on a fuzzy problem but it can also render unhealthy framing of the problem scope. HCI, as a research topic, attracts an opportunist lot of researchers and there is no topic that, with or without mild effort, is deemed unfit for the community. Every activity that could involve a computer is relevant to study, from any imaginable perspective.

Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them (ACM 1992b).

I do not oppose this, but it has led to an explosion of directions for research. At the latest CHI conference,\(^1\) nine different themes were especially promoted: Design, Engineering, Management, User Experience, Child-Computer Interaction, Digital Arts, Games and Entertainment, Health, and Sustainability. As computers nowadays are ubiquitous, the interaction between humans and computers is interesting for researchers from a wide variety of disciplines. This type of agnosticism has both positive and negative effects. Among the positive outcomes is a curiosity for new ideas and thus a fast adoption of new insights. Among the negative, we can find something which I think of as close to institutionalized charlatanry, where rigid concepts are borrowed from other disciplines and translated into actionable heuristics with weak theoretical justification.

\(^1\)CHI is the largest publication outlet for HCI research: http://chi2013.acm.org

Figure 5.1: HCI in 1992, as described in ACM SIGCHI Curricula for Human-Computer Interaction. Source: ACM (1992b)
Figure 5.1 has been used many times to introduce the area of Human Computer Interaction (HCI). Perhaps even too many times, as it has started to live a life outside of its context in the Curricula\textsuperscript{2}.

Indeed it visualizes well the scope of the field at the time of its formation but without the supporting manifest it leaves to the viewer or presenter to imagine how the various topics are addressed and, perhaps more importantly, not addressed within HCI. As I curiously went to the source (ACM 1992b) to become enlightened about these constraints, I found that practically any research that is involving humans and computers matches the description. In fact, what is today often regarded as neighboring research areas, like Information Systems, Computer Graphics, Artificial Intelligence, and Operations Research, are also included. The first version of the document dates back to 1992, so it is understandable that the approach was inclusive rather than exclusive, as the impact of computers was still comparatively limited. However, as many researchers today have a somewhat narrower definition of what the area covers, we ought to find a more descriptive definition or start thinking in new terms.

To conclude this somewhat unfair account, I suggest that we can expect the discipline to very soon be subdivided into several more strictly defined topics, simply because it is difficult, or even impossible, to survey—not to mention scrutinize—all the ideas that are currently put forth. I am anticipating the situation when every research department, in every research discipline that has something to do with humans, has a group that is analyzing computer use from their respective theoretical point of view.

5.1.1 The State of HCI

Still ten years ago, in the foreword to the Human-Computer Interaction Handbook (Jacko and Sears 2003), Ben Shneiderman uses the various phases of an individual’s development as a metaphor to illustrate how the field of HCI is continuing to grow and develop. He was not sure how far the development has reached: is the field in its childhood, or has it perhaps reached its teens or even become a grown up, yet not fully matured? Further, he describes three different options for the field to develop academically:

It could mature into a full-fledged basic discipline, such as physics or psychology; it could continue to be a inter- and multidisciplinary meeting place; and it could grow into a practice-oriented discipline, such as architecture or medicine. These reflections illustrate something that I consider as a problem within the area—namely the reluctance, or inability, to clearly define the boundaries and the purpose of research within it.

\textsuperscript{2}ACM (Association for Computing Machinery) SIGCHI (Special Interest Group on Computer Human Interaction) Curricula for HCI (ACM 1992b).
5.1.2 The Future of HCI

Personally, I have become comfortable with the idea of HCI as an interdisciplinary research topic and not as much as a discipline. Within this topic, there is still need for a discipline to study interaction design, with researchers who use knowledge generated in other disciplines to inform design. I hope that psychologists, sociologists, and ethnologists will start doing more HCI research within their core disciplines to complement those who have migrated to the discipline of HCI. This would sustain methodological rigidity and would counter the tendency of creating HCI specific strands of other research disciplines.

On the other hand, if HCI is to become recognized as a basic discipline, HCI researchers need to somehow settle for a clearly defining agenda. Humans interacting with computers is not distinct enough as demarcation anymore. One idea comes from Gilbert Cockton (Cockton 2004). He mentions a bit in passing something that probably would not unite the plethora of HCI researchers but nevertheless could create a stronger identity in the community: in contrast to other disciplines that are interested in the meeting place between humans and computers, e.g., sociology, psychology, economics, and media studies, HCI researchers have a primary objective to actually improve interaction design. After some reflection, this idealistic pledge starts to sound pretty concrete.

The simple definition sets clear boundaries for what is, and what is not, HCI research. In Cockton’s terminology it means that we should, based on an understanding of how computer use affects human values, derive concrete objectives to support fulfillment of these values. This means that it is neither enough for HCI researchers to satisfy with ethnographic studies to observe the current situation, nor psychological experiments to determine some cognitive capacity (this is probably done better by “native” ethnologists and psychologists), and it is not enough to evaluate and design systems using existing methods (this is what practitioners should do). HCI needs researchers with diverse perspectives, united by an interest to understand, explain, and improve (according to the criteria that are currently in fashion) interaction with computers.

An HCI researcher should not have to be a polymath. She should be a confident user of knowledge produced in other disciplines, without necessarily being an expert in how this knowledge was produced. As computers are becoming ubiquitous and, consequently, a widespread research interest, HCI serves an important role as medium between research on humans that is involving computers, and research on technology. HCI researchers have the perspective, methods, and knowledge to understand how technology affects human values, and to understand that technology should be shaped to accord to these and not the other way around.
The dominant research agenda within HCI has changed over time—from computer systems thinking, over cognitive psychology, to ethnography and contextualization (with a brief visit in something that in retrospect probably will qualify as the decade of hedonism and value nihilist user experience). I agree with Cockton’s proposal to redefine the field to focus on basic human values that justify the use of computers. With this overarching perspective, all the previous hegemonies could contribute with methods and theories to better understand what users both want and need.

It may sound overly demanding that researchers in a relatively young research area need to come to agreement on the research agenda, particularly since the object of study has undergone an unprecedented development during its establishment but, in order to gain legitimacy, it might be necessary to be very concrete about what HCI research focuses on. I suggest that the definition from the ACM SIGCHI Curricula (ACM 1992b) has served its purpose well, but that it is time to replace it with one that anchors in what it is to be human.

5.1.3 The Human in HCI

Much of HCI research has focused on details about computer interfaces, socio-technical systems and even psychologically relevant aspects like perception and cognition. Much less attention has been paid to the fundamental philosophical goals of interacting with computers. Most HCI researchers would agree that the human is the important component in the relationship, but for that to be credible, there should be a proper idea about what it means to be human. We surely have a common-sense understanding about it, as we are observing humans every day in our lives, but can we take for granted that what we observe really captures the essence of being human?

I think it is time for HCI researchers to start reconsidering the scope of what to take into account when determining what is a good artifact and what is good usability. Neither the usability, nor the user experience (Hassenzahl and Tractinsky 2006) definitions take into account what happens with people when they use technology. Usability is often interpreted as suitability for a task and ease of use, which is not a terrible simplification of the way the term is defined. This appears to be a very desirable feature of technology, but it can lead astray.

Consider, for instance, the rice cooker in Figure 5.2. Apart from the obvious inconsistency, that the display is in Japanese and the labels are in English, the machine is very easy to use for its intended purpose. The 3-step procedure to make perfect rice is:

1. Measure rice with a dedicated cup—one cup equals one serving of rice. Optionally you may wash the rice first. According to experts, you should.
2. Fill up water to the marker inside the basin that corresponds to the number of cups and the type of rice.
3. Close the lid and push COOK.

If you are using special kinds of rice, you should choose the corresponding program and amount of water. That is all you need to worry about.

Now, this may seem like a nice piece of kitchen appliance. It is efficient, effective and does not cause much discomfort for the user. From a usability perspective, it is not bad. The user experience of it is more difficult to assess. The cooker is in my personal, subjective opinion not aesthetically appealing but then again, I would probably not pay more for a rice cooker that could double as a bibelot. More relevant would be to assess the quality of the resulting rice, which is excellent. Also, the appliance minimizes my cognitive workload, as I do not have to remember when to take the rice pot off the stove; the rice cooker will signal when the rice is done—recognition rather than recall.

Still I have one, in my opinion important, objection toward unreflected adoption of this kind of usable devices: the rice cooker prevents me from learning the skill of boiling rice. I do not learn which temperature to use, I do not learn for how long I should boil it, I do not learn the proper ratio between rice and water. All I learn is to do the work that the machine is unable to perform: fill up rice and water and decide when I want to prepare my meal. To most avid rice eaters, this particular skill may not be a great sacrifice in face of the convenience that the rice cooker offers. If there is always one available, it is an unnecessary skill to be able to cook it in just any pot. I will not contend this. Still, it is designed for a single purpose and most people will use it only for that. A few Google queries will, consequently, suggest that more people are fascinated about what else, besides rice, can be cooked in a rice cooker.
than about what can be cooked in a frying pan. It seems like this idea is something worth writing about; it has news value as it is not familiar. I see this as an example of functional fixedness, a concept that I will detail later, but it is also an example of when I have transferred some of my autonomy to a device.

I cannot shake off this itching suspicion that usability norms up until now have done harm along with the goods that they have inspired. Take for instance the ISO definition of usability above. My impression has for years been that it is very sensible. Who could object against the values and the sound delimitation of the problem scope? For a long time I even used the definition to argue for the relevance of my research area, but I am not as confident in the soundness of it anymore. Why are efficiency and effectiveness so important that they deserve to define usability? And satisfaction sounds mostly like a mockery of what drives people to do a good job: *if workers are satisfied they do a good job?* To illustrate my point, I would like to offer a different, hypothetical definition, with different connotations before I proceed.

Usability is the extent to which a tool allows deliberate actions by a sufficiently trained user.

In this definition, actions should be understood in the same way as in the previous definition of tools, as purposefully performed procedures. Like the tool definition, it does not assume that we, as humans, are heading somewhere; that there is an ulterior goal that can be subdivided into smaller goals, until we have something that can be formulated as a task. It assumes neither activities nor goals. It does, however, assume agency from the user and it focuses only on how well an artifact, as a tool, can be used for this purpose.

5.2 Sketchpad—A Man-Machine Graphical Communication System

To illustrate a point about HCI, functional fixedness, and computers as tools, allow me to make a brief historical remark.

Sketchpad was a software interface for the Lincoln TX-2 computer, created by Ivan Sutherland (1963) during his PhD studies, with the main purpose of creating interactive drawings. Sutherland himself presented the tool as a “man-machine graphical communication system”, which is interesting from a semantic point of view. Did he consider the computer as some kind of companion? And did he consider the creation of a drawing as some kind of negotiation between the user and the computer?

The Sketchpad system uses drawing as a novel *communication medium* for a computer. The system contains input, output, and computation programs which enable it to interpret information drawn directly on a computer display. It has
been used to draw electrical, mechanical, scientific, mathematical, and animated drawings; it is a general purpose system. (Sutherland 2003, abstract, stress added)

The Sketchpad was created using a recently developed display technology with a light pen connected to the display in a way so that the user could directly manipulate objects on the screen (see Stotz 1963 for a technical description).

There are clear indications in Sutherland’s thesis that part of the design was deriving from the available technology. Perhaps the most striking example is the description of how the chosen format for storing drawings led to one of the most interesting functionalities of the system, namely the recursive merging that facilitates the object-oriented approach that the system is now famous for.

The particular form of the ring structure chosen has led to some of the most interesting features of the system simply because the changes required to keep the ring structure consistent led to useful facilities such as recursive merging. (Sutherland 2003, p. 34)

Technology seemed to have a strong impact on the actual design of the system, but what were Sutherland’s incentives to create the system? The following quote gives a clue:

The decision actually to implement a drawing system reflected our feeling that knowledge of the facilities which would prove useful could only be obtained by actually trying them. The decision actually to implement a drawing system did not mean, however, that brute force techniques were to be used to computerize ordinary drafting tools; it was implicit in the research nature of the work that simple new facilities should be discovered which, when implemented, should be useful in a wide range of applications, preferably including some unforeseen [sic] ones.

[…] Had a working system not been developed, our thinking would have been too strongly influenced by a lifetime of drawing on paper to discover many of the useful services that the computer can provide. (Sutherland 2003, p. 8, stress added)

It seems like Sutherland was the type of person who did not satisfy with theorizing about the usefulness of a proposed system, but wanted to explore them in use. Furthermore, he is hypothesizing that ingrained habits tend to reduce the possibility to imagine something different. This is an interesting attitude in light of the current-day paradigm within HCI, in which the users’ immediate needs are seen as the correct starting point for development. Psychologically, there is definitely merit in Sutherland’s line of reasoning. He seems very aware that introduction of new technology will change the way people work and think. Noteworthy is also the description of his innovations as “simple new facilities” that could later perhaps become useful.
Sutherland invented several novel technologies in the application, among which are the interactive graphical user interface, the event-driven programming language, and an object-oriented approach to creating the drawings. Alan Kay (1987) reports in a video to have asked how this was possible to achieve in just one year, to which Sutherland had replied: “I didn’t know it was hard!”

This is actually very interesting for the idea in this thesis. At the time when Sketchpad was created, there were no precedents and no established patterns for how to construct this kind of system. What Sutherland had to work with was a problem, and a machine that could be configured to address that problem. The main source of influence on his work was probably the way that engineers were drawing on paper, and the frustrations that these had in their work.

As nobody had any preconceptions about how the system should be constructed, the degrees of freedom for designing a purposeful system were many, and constrained only by the available technology. In fact, when Sutherland presents the usefulness of the Sketchpad interface, he brings up advantages that are resulting from classic software functionality rather than from the interaction design: the possibility to store and update drawings, the propagation of changes from the master drawing to instances of it, the possibility to reuse the relations between objects in simulations, and the possibility to copy objects recursively (Sutherland 2003). It seems like Sutherland considered the user interface as a natural result of the preconditions; as brought forth through poiēsis (Heidegger 1977). As this was one of the very first user interfaces, a separation between the user interface and its technological implementation was not yet established.

Semantically, there is an interesting progression in the use of the word “tool” in Sutherland’s thesis. In the beginning he is only using it to refer to craftsmen’s drafting tools, and to a programming language for machine tools called Automatic Programming Tool. Sketchpad is up until the end something that is vaguely described as a system. Finally, at the end of the thesis, all hesitation is gone and he acknowledges both the interface and the whole Sketchpad system as a tool.

With Sketchpad we have a powerful graphical input tool. [...] What we do claim is that a graphical input coupled to some kind of computation which is in turn coupled to graphical output is a truly powerful tool for education and design. (Sutherland 2003, p. 129)

To me this is interesting, as it suggests that a linkage has been established between the traditional tools of craftsmen and the computer system. There, however, seems to be some kind of ambivalence in how the system should be addressed. I lack the contextual knowledge to judge if the terminology is accidental, or if a communication medium actually was considered to be a more
specific description of an input tool. A likely explanation is that Sutherland used the term tool in the same way as I have defined, as arising in purposeful use.

However, what is most important for the remainder of this thesis is the process that lead to the development of Sketchpad. By using the properties of available technology to address a problem that he saw in current practice, he was probing into the future; he created an artifact as a hypothesis.

5.3 Ethical Interaction Design

Within HCI, ethics have this far played a rather peripheral role. As many other writers of textbooks who are directed toward a professional audience, Cooper, Reimann, and Cronin (2007) reveal little interest in what really is ethical conduct. In an approximately one page long paragraph (pp. 152–153) in their 600+ pages book *About Face 3*, they name two topics of ethical import: “Do No Harm” and “Improve Human Situations”. Under the first heading they have gathered a few types of harm that can be done: 1) interpersonal, 2) psychological, 3) physical, 4) environmental, and 5) social and societal. Of these I find the brief descriptions of 2 and 5 most in need of further discussion. Psychological harm means in their words: “confusion, discomfort, frustration, coercion, boredom”. Social and societal harm means: “exploitation, creation, or perpetuation of injustice”. There are no further explanations. I do not take this as an exhaustive account of how they understand ethics, but it is rather curious that a profession that is deeply affecting human activity does not display higher demands on ethical reflection. Practical experience yields incredibly valuable advice, but without being explicit about what the premises are, you may also deliver advice that is detrimental to your cause. In a sense, this is a reflection of our society. We do not talk about ethics, other than in vague blurbs about good and right, and most of the time it works out fine anyway. In fact, I believe that it is rather common to become a bit uncertain when it comes to what is right and wrong on a philosophical level. Perhaps philosophers have been bad at inviting non-philosophers to the dialogue.

Value-Sensitive Design

Probably the most known approach to ethics within HCI is Value-Sensitive Design (B. Friedman and Kahn Jr. 2003; B. Friedman, Kahn Jr., and Borning 2008). It is described by the authors as an interactional theory to take into account values in design. By interactional, they mean that values regarding technology are determined by the goals of its user, and neither inscribed in the technology, nor determined by social forces.

[…] the interactional position holds that while the features or properties that people design into technologies more readily support certain values and hinder
others, the technology’s actual use depends on the goals of the people interacting with it. (B. Friedman, Kahn Jr., and Borning 2008, p. 86)

From how it is described, I understand the interactional label to imply that technology is essentially value-neutral. To determine how to design technology to address values in use, the framework offers a range of values, sampled from several different sources. An analysis that is focused on any set of these values is suggested, in order to determine whether there are any risks for conflicts due to the use of technology. This is performed through “conceptual, empirical, and technical investigations, applied iteratively and integratively” (B. Friedman, Kahn Jr., and Borning 2008, p. 86), and it “distinguishes between usability and human values with ethical import” (B. Friedman, Kahn Jr., and Borning 2008, p. 85) since usability is considered as a property of technology, and not a human value like the following:

[...] Value-Sensitive Design enlarges the scope of human values beyond those of cooperation (CSCW) and participation and democracy (Participatory Design) to include all values, especially those with moral import. By moral, we refer to issues that pertain to fairness, justice, human welfare, and virtue, encompassing within moral philosophical theory deontology (Dworkin, 1978; Gewirth, 1978; Kant, 1785/1964; Rawls, 1971), consequentialism (Smart and Williams, 1973; see Scheffler, 1982 for an analysis), and virtue (Campbell and Christopher, 1996; Foot, 1978; MacIntyre, 1984). Value Sensitive Design also accounts for conventions (e.g., standardization of protocols) and personal values (e.g., color preferences within a graphical user interface). (B. Friedman, Kahn Jr., and Borning 2008, p. 85)

The value in Value-Sensitive Design, as I see it, is that it offers opportunities to discuss important matters, but it does not offer much guidance for how to design something in a more ethical way. For that, the framework is too generous. Later in this thesis I offer a more normative approach to designing and analyzing artifacts, by addressing the universal, abstract value of autonomy:

Generally, the more concretely (act-based) one conceptualizes a value, the more one will be led to recognize cultural variation; conversely, the more abstractly one conceptualizes a value, the more one will be led to recognize universals. (B. Friedman, Kahn Jr., and Borning 2008, p. 86)

Even though ethics is rarely mentioned explicitly within HCI, it is embedded in other choices and practices without necessarily being acknowledged. Only when something seems terribly wrong will we notice that there was an ethical dimension to it. In this thesis, I want to give tools for thinking more deliberately about what is moral. Ethical deliberation should not be something that is attached to an analysis, but should be an integral part of good praxis (Bannon 2011). There are no value-free practices.
Part II:  
Thinking Tools, a.k.a Theory

In this part I discuss the theoretical concepts that guide my reasoning. It contains a curious set of concepts from other thinkers as well as my attempt to synthetize these into something of a coherent theory.
6. The Human Mind

I will account for a minimal set of psychological concepts that I consider to be important for understanding how to design artifacts for people. It will hopefully suffice to understand the essence of this thesis. Even though I will give a rather soul-less account of the human mind, I do not want to imply that we all work in the exact same way. We are all individuals with our respective quirks. Even in the case that we found ourselves to be more alike than we actually are, we would still make sure to perceive and make a fuzz about what makes us different from them. As humans we are very alike but we appreciate the small variations that exist and use and abuse them to distinguish ourselves.

6.1 Mental Models

The first concept that I wish to introduce, is the mental model. In one sentence, it is the understanding that a person has about how a system, in the word’s most general meaning, works. The construct has been studied extensively in cognitive science but has since proven useful in HCI and the more safety oriented field of Human Factors Engineering. The term mental model suggests an ordered system with some specific processes going on in the brain. This is not necessarily a true representation, but the construct is useful to facilitate dialogue about understanding and how to design with this in mind. In a sense, to an observer the mental model is a black box that takes input and gives output.

A commonly used example of mental models at work is to ask people how they understand electricity, this fascinating and mostly invisible phenomenon (D. Gentner and D.R. Gentner 1983). Many understand it to behave metaphorically like water: Plug in your apparatus into a wall outlet and electricity will “flow” into it, like a stream. Current is the width of the stream, voltage is the water pressure. This is not an accurate metaphor but for many purposes it suffices. It is easy to imagine that high current is dangerous while high voltage is not, if the current is low. For other purposes, it will, however, create obstacles for proper understanding. I will not go into the details here but only conclude that mental models are usually formed by experiences, both own and other’s, with the purpose to give a sense of understanding of what is happening and thus be able to predict outcomes of actions. Ironically, different pundits detail this concept differently. They have, so to speak, different mental models of what mental models are.
Donald Norman summarizes six features of mental models that are important to remember in systems design:

1. Mental models are incomplete.
2. People’s ability to “run” their models is severely limited.
3. Mental models are unstable: People forget the details of the system they are using [...].
4. Mental models do not have firm boundaries: similar devices and operations get confused with one another.
5. Mental models are “unscientific”: People maintain “superstitious” behavior patterns even when they are unneeded because they cost little in physical effort and save mental effort.
6. Mental models are parsimonious: Often people do extra physical operations rather than the mental planning that would allow them to avoid those actions [...]

(Norman 1983, p. 8)

For my application, it suffices to realize that the user’s understanding of how an artifact works may be more or less precise, flawed, or incomplete, and that many rather do extra work than reflect over whether they have a a flawed understanding. This means that different people may interpret the same observed phenomenon differently. To counter this, should not artifacts be designed so that they support the formation of an appropriate\(^1\) mental model? Not everybody agrees with this position.

Alan Cooper, the practitioner that has written several influential books about interaction design, claims in his latest, About Face 3: The Essentials of Interaction Design (co-written with Robert Reimann and Dave Cronin) that

One of the most important goals of the designer should be to make the represented model [what the user can observe of an artifact] match the mental model of users as closely as possible.” (Cooper, Reimann, and Cronin 2007, p. 30)

This approach assumes that the user has a mental model of a system that is suitable for the task or activity that she is to perform. In contrast, rigorous processes like Cognitive Work Analysis, here represented by Kim Vicente, warns explicitly for the fact that users’ mental models may not always be correct:

If we base the design of an interface on an analysis of workers’ current mental models, then we will reinforce any existing misconceptions that workers already have. (Vicente 1999, pp. 52–53)

These two pieces of advice are clearly contradictory so we should understand why. My best guess is that they were thinking about different audiences when

\(^1\) I refrain from writing “correct” as it is epistemologically impossible to establish what a correct mental model is.
they made these claims. For Cooper et al., it is important to create artifacts that are easy to use, while for Vicente it is important to create artifacts that give the user as much control as possible, as he is targeting, among others, users who are working in safety critical settings. One is tempted to ask a blunt question: Is this a battle between design for dummies and design for professionals? Such an interpretation can be further reinforced by continued reading of Cooper et al:

\[
\text{[I]f we create represented models that are simpler than the actual implementation model, we help the user achieve a better understanding. […] One of the most significant ways in which computers can assist human beings is by putting a simple face on complex processes and situations. (Cooper, Reimann, and Cronin 2007, p. 30)}
\]

Now the rhetorical question is: What does Cooper et al. mean by understanding? A well-intentioned reading would suggest that what he means is that users should not have to form a mental model of the operating system and the algorithms that are used to run the artifact, because these insignificant details will not help her use the artifact for her purpose. A more critical reading suggests that he is confusing matters. I do not claim that interface details have to be mirror images of the respective system functions but they should correspond distinctly to one. The interface should be an abstraction, not a distortion of the system model. Colors in photo editing software should be possible to also choose visually rather than only by typing hexadecimal codes.\(^2\) Search queries should not require the user to be fluent in boolean algebra and it is definitely not a good idea to design a website with a dedicated search box for each field in the database. However, it should be obvious to the user that she has searched in every field, if that is the case.

Recently a Swedish journalist was shamed for having confused news reports about a place with reports about an artist and a race horse (Bjereld 2013). She had assumed that a query for the name of the place would render only articles about that place and not its namesakes. This particular case probably did not involve a database with separate fields but it highlights the risk of supplying a system that makes something easy to do, which is not necessarily that straightforward. The system did not inform her of what it had done, it just returned a convenient, false, number. Always when the system model is hidden from the user will there be a risk of tricking the user into drawing the wrong conclusion.

An appropriate mental model is necessary if the user is to maintain her autonomy. To be able to form this, design should communicate an appropriate abstraction of its system model. By appropriate abstraction, I mean that details that can be guaranteed to work predictably—that the user does not need to observe in order to form a mental model—can be abstracted. If this is

\(^2\)It is useful for, e.g., web designers to have both alternatives available.
impossible—if the system model is too complex—the artifact should not be salvaged by a layer that hides the mess from the user. Such an artifact has been developed in the wrong order, under the assumption that a superficial user interface can be added afterwards.

One of the common positions that I challenge here is that the user interface should be kept separate from the system model. I agree with this in technical terms—the MVC design pattern is sound, and user interface code should not be mixed up with code for functionality—but it seems that the justification for this has become confused so that the functionality of a system is seen as disconnected from its presentation.

6.1.1 The User’s Mental Model

Users’ mental models are relevant in HCI for two purposes. Either 1) to address the problem of designing an artifact so that the user can form a proper understanding of its function, or 2) to address the problem of figuring out the user’s understanding of relevant relations and mechanisms in her task, and design an artifact accordingly.

Both of these problem formulations focus on the user and in neither of them is it obvious how the designers and developers come to understand exactly what to construct. This is often assumed to be a problem of information transfer rather than cognition and, consequently, the advocated strategy is that the designer should inquire as much as possible about the intended users.

In the first problem formulation (1) there is an assumption that the user will form specific mental models for specific artifacts. The task of the designer thus becomes to design an artifact so that it communicates how it works—in domain terminology: its system model or representational model—so that the user can form an appropriate mental model of it. This is often a good approach, even though it is difficult to design an artifact of some complexity so that it becomes reasonable to grasp its model. More importantly, it will lead to problems if the designer has not understood what the user really needed or wished for in the first place. The lesson to learn from this has been made into a slogan that is something like: it is more important to design the right system than to design the system right (Buxton 2007; Greenberg and Buxton 2008).

The second problem formulation (2) is often advocated when designing for users who do not wish to become experts in using your artifact, with the implication that they only need a basic, functional understanding of how it works. Web design is a good example of this. On the web, people usually want to find information or get a task done with least cognitive effort. Since visitors to your website spend most of their time on other sites, it is wise to align to established standards, as those will have influenced the visitors’ expectations of how things work. It is thus sensible, for instance, to place information where
visitors look for it instead of maintaining your own idiosyncratic structuring, even if that would make better sense from the perspective of your enterprise.

This approach has good applications, but if applied too frivolously it also risks reducing the possibility for a user to form a purposeful mental model of how an artifact is involved in the task that is to be performed. If details are handled as particular—which is the result of conforming to standards regarding particular functions—it can become difficult for a user to form a coherent mental model of the artifact as a whole. Consider as an example the application Evernote. It is a note-taking tool that allows the user to quickly make notes, which are automatically saved and synchronized between different devices. A note made on the smartphone swiftly appears on the computer and vice versa. Contrary to the fact that the user does not need to save notes, the computer application offers a save function, both as a menu option and as a keyboard shortcut. Why? The answer is probably that some users did not feel certain that their notes really were saved, and to relieve their anxiety, the designers added a dummy function, that in reality does not do anything apart from disturbing the user’s possibility to grasp what is happening in the tool. An unnecessary ritual is introduced.

6.1.2 The Designer’s Mental Model

Although standards are good, as they support recognition and by that lowers the barrier for learning, they may become an obstacle to optimal solutions. This is embedded in the concept of a standard as the best compromise, and for many applications the value of compatibility is more important than perfect fit for a task. When we are dealing with tools, we are in the domain of qualitative judgment, where a balance needs to be struck between ease of learning to use it, compatibility with other tools, efficiency, and not least: the user’s autonomy.

Some tools are very difficult to use without fully understanding their mechanisms, but they can give great powers to a skilled user; the text editor Vim is a close candidate. Other tools are familiar and easy to learn but sometimes do misdirected magic that the user cannot recover easily from; the text editor Word is more in this direction. It is usually not desirable to maximize one aspect on behalf of others, but what I argue is that the balance has to be determined by the user’s purpose, which is what the designer needs to identify.

In current literature, it is usually taken for granted that the designer of an artifact is able to form a correct mental model, if she is given enough information about the user. Donald Norman almost address this when he writes about Conceptual Models.

Conceptual models are devised as tools for the understanding or teaching of physical systems. Mental models are what people really have in their heads and what guides their use of things. Ideally, there ought to be a direct and
simple relationship between the conceptual and the mental model. All too often, however, this is not the case. […] all too often there is no correspondence among the conceptual model of the system that guided the designer, the system image that is presented to the user, the material that is presented in instructional manuals that is taught to the user, and the mental models of the user. (Norman 1983, pp. 12–13)

I do not doubt that Norman understood that it is important for designers to form a good understanding of the user and her situation, but many interpretations of his thinking assume that designers should communicate a conceptual model through design. This is fine, but if designers’ conceptual model is based on flawed mental models of the users’ situation, it does not help much if this is better communicated. The wrong system will just be better designed.

Participatory design (Ehn 1988; Gulliksen et al. 2003), the school of user-centered design that originated in Scandinavia, solves this problem by involving users as co-designers. The original incentive is egalitarian, but the effects are not limited to a political value. Having users as co-designers means that designers, at least in theory, have access to users’ mental models of their work. Regrettably, one user’s model may not be same as the others’, and users quickly adjust their mental models as they reflect over their work tasks and learn more about the system model. They thus become to view the system more from the point of view of a designer rather than a user: “Involving users on a full-time basis in a project quickly turns them into domain experts rather than representative users” (Gulliksen et al. 2003, p. 407). This can be countered by awareness, but a more critical obstacle to wide adoption of this design practice is that it requires serious dedication from both designers and users.

Iterative, user-centered design (Gould and Lewis 1985) is by some claimed to be possible to perform without extensive engagement. User-centeredness does not necessarily require users to be co-designers, but it is stressed that regular user contact is necessary in order for designers to understand what to construct. This contact can mean, for instance, interviews and observations. What I find striking in this approach is that designers’ mental models are assumed to naturally become refined as they are introduced to the users. This is an epistemological assumption that I do not share. I do not see understanding as the natural consequence of being exposed to information. In this thesis I try to devise a way for designers to form a proper mental model of what users need, so that they can design the right tool or artifact. When they know what this is, their expertise can be utilized to get the design right.

6.2 System I and System II Thinking
There is almost consensus in psychological literature that there are two fundamentally different cognitive processes involved in decision making. Many
philosophers tend to agree, although their focus is more on the method of choice than the cognitive mechanisms affecting it. I will discuss the difference a bit later.

The simplest of the cognitive processes is what some psychologists call intuitive judgment, associative or just System I thinking. It is the automatic, instant responses to stimuli that relieve us from mental computation and thus constitute a very economic strategy. The more complex one is sometimes called rule-based or System II thinking (Sloman 1996; Stanovich and West 2000; Sunstein 2005; Kahneman 2012).

I use these concepts to illustrate the observed difference between how we handle cognitive tasks. Despite using the terms System I and System II, I do not take a side in the physiological question whether there actually are two types of cognitive processes, or just one that is employed to various degrees. (Stanovich and West 2000).

6.2.1 Heuristics and Cognitive Biases

Tversky, Kahneman and other cognitive psychologists fought long against the belief that humans are making rational decisions (Tversky and Kahneman 1974; Tversky and Kahneman 1983). By carefully designed experiments, they have shown that when humans make decisions, we tend to both neglect information and invent information that is not really given, just inferred from compelling formulations. Within psychology, this phenomenon is known under the term cognitive heuristics, which are systematic and useful rules of thumb. In order to be able to cope with the complexity created by a multitude of interactive everyday events, it is necessary for us to make predictions and assumptions about the likely outcomes of decisions. Heuristics belong to the System I processes and are very useful because they require less effort to handle information.

In most cases they serve us well, and only in certain situations do they cause fallacies. Many of these fallacies have been identified in isolation and subsequently earned a name. Daniel Kahneman was awarded The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel in 2002 for the work he did in this field together with the late Amos Tversky (Kahneman 2003).

On the one hand, it is probably good that decision makers have started to realize that they are biased by irrelevant details, like how information is phrased—they will perhaps become a bit more careful—but on the other hand, there may be a risk also in focusing too much on the negative side of biases. If one accepts the theory that we have evolved by natural selection, then cognitive biases probably provide a competitive edge, compared to unbiased thinking. By trusting our intuitions we can make quick decisions, and when these fail, they fail on the side of what is likely rather than on what is unlikely.
However, as these biases sometimes become an obstacle in the evaluation of information, I below summarize a sample of well-known biases that are relevant for the present thesis.

**Overconfidence and Confirmation Bias**
Overconfidence is an unfounded exaggerated belief in one’s own ability and knowledge (Kruger and Dunning 1999). Confirmation bias means to search for evidence that confirm one’s belief and being reluctant to accept evidence that points in other directions. This also means that initial impressions will frame our understanding of subsequent information. According to Kahneman, this happens easily in situations in which it is difficult to form a proper mental model, and we have to base judgments on limited information (Kahneman 2012).

**Anchoring Bias and Base-Rate Fallacy**
An anchoring bias means that evaluation of information is anchored in some, possibly arbitrarily chosen, assumption or piece of information. The question “was Gandhi more or less than 144 years old when he died?” will nudge people into thinking that he was older than he in fact was (Kahneman 2012). I will below discuss this as a problem in the approach that I advocate.

The base-rate fallacy means that when evaluating the likelihood of an event, it is easy to forget that there is a base rate that affects how probable it is. Inexperienced physicians may find that the symptoms of a patient exactly matches those that indicate an exotic disease; the evidence seem overwhelming, but in reality it is much more likely that the symptoms are created by two separate, common diseases. The statistical probability for the latter is much higher but this is forgotten in light of the perfect, yet improbable, match.

**Einstellung Effect or Maslow’s Hammer**
Maslow’s hammer is the popular name for the tendency to solve every problem by using the same, familiar tool. “If you have a hammer, then everything looks like a nail”. Einstellung effect is a similar phenomenon. For instance: if you are solving several problems by using one method, it is likely that you will attempt to solve also the next problem using the same method, even if it is impossible (Anderson 2010).

The generalization of this is that we attempt to initially interpret phenomena by using a familiar explanation. From a meta-perspective, this is highly relevant for this thesis as I am using the principle of autonomy to explain several observations. On the topic level, it is important for designers to be aware of the fact that people tend to favor doing things in the way that they are used to.
6.2.2 Functional Fixedness

Functional fixedness is closely related to the einstellung effect and Maslow’s hammer. When we have become accustomed to a thing and its intended functionality, to the extent that it becomes difficult for us to imagine other uses and functions for this thing, we are experiencing something that is called functional fixedness. The expression is related to the term affordance, coined by James Gibson and popularized by Donald Norman\(^3\) in *The Design of Everyday Things* (Norman 2002).

Gibson used the term affordance to describe the different ways in which an object is possible to use or interact with for a specified individual. E.g., for me a small stone affords grasping, lifting, throwing, caressing, rolling, etc, whereas it for an ant affords climbing. The term has a direct derivative in perceived affordance, i.e., the ways in which a person understands that an object can be interacted with. Functional fixedness is a limited set of perceived affordances in relation to actual affordances, resulting in a mental model in which the actual affordances of an object have been replaced with specific, familiar functions.

A classic experiment from 1931 illustrates this very well. Briefly summarized (see Figure 6.1), test subjects in the experiment were asked to come up with four different solutions to a problem, which was to tie together the ends of two strings hanging from the ceiling of a room. The strings were too far apart to make it possible for the test subjects to take one string and walk over to the other to tie them but in the room there were also several types of objects to aid in the task. Among other things, there were poles, pliers, chairs, and extension cords. Three of the solutions did not presuppose the test subject to change model of the strings hanging from the ceiling, and were generally easy to find (extending one of the strings, pulling one string with the pole, and tying one string to a chair and moving it closer to the other string). The fourth solution, however, required them to realize that one of the strings could also function as a pendulum, given that a weight is tied to the end of the string. Only a minority (39.3%) accomplished this task within 10 minutes, after which the observer intervened. The remaining were given tacit stimulation; first by the observer briefly swinging one of the strings while passing it, and if that was unsuccessful by suggesting a pair of pliers as the only necessary tool to complete the task. Both of these hints were done in order to help the test subjects adopt a new model of the string. After this an additional 37.7% completed the task (Maier 1931).

\(^3\)Norman has taken a lot of heat because of how he was using the terms. Most readers understood him to claim that design should convey perceived affordance, by adding cues to intended functionality, which is different from the way Gibson intended the term to be used. Norman has since changed position and nowadays prefers to use the term *signifier* to describe signs that convey information about function or state, whether these are intentionally created or not (Norman 1999; Norman 2008).
Figure 6.1: An example of a functional fixedness problem. What qualitatively different means does the environment present to reach both of the two strings so that they can be tied together? There were plenty of artifacts in the room. The significant ones are in the illustration.

This experiment, as other subsequent versions of it (e.g., the one in Figure 6.2), shows that many people have difficulties to perceive something that has a certain gestalt (i.e. a “whole”) as having also other gestalts. We usually do not think about the actual properties of objects but instead chunk them into familiar units. Furthermore, in the experiment there was a strong priming to regard the strings in one specific way, as the three more obvious solutions were based on this model. The string was perceived as a passive object—a target—which made it difficult for the test subjects to change their model of it to a tool that could help them accomplish the task.

This phenomenon has many repercussions on tool use in general. The tool definitions that I sketched above are in a sense idealistic. In everyday life we do not think about the familiar objects around us as potential tools, unless we have been accustomed to do so. And for tools that have an established intended use, like a hair dryer, it is difficult to imagine other uses, e.g., for heating food, especially if these other uses require us to re-imagine the function of the tool. Anecdotal evidence suggests that novel uses for objects often are discovered by chance. Among many others, the discoveries of radioactivity, penicillin, Kevlar, Teflon, Saccharin, etc. seem to support such a conception. These are
stories of obvious chance but less dramatic serendipitous discoveries are likely more common than most would like to admit.

The story of Sultan, the clever chimpanzee, sustains this hypothesis. Sultan became famous after proving problem-solving skills in front of his observer Wolfgang Köhler, a German psychologist notable for being one of the founders of Gestalt psychology. In one of the challenges that Sultan was forced to go through, a bunch of bananas was placed outside of his cage and he was supplied with two sticks, neither of which was long enough to reach the bananas so that he could pull these into the cage. However, the sticks were hollow and chosen so that one of them fit into the other one with the result that a long enough rod could be created. Köhler wanted to know if Sultan would be able to figure out this rather advanced assembling of the tool. He was not. Not even after observing Köhler put his finger into the hollow of a stick could he solve the puzzle. However, after playing around with the sticks for some hours, and during the play accidentally happening to join the two sticks, he immediately recognized the solution. In subsequent repetitions the task was easy for him. What this suggests is not only confined to apes. Also humans are good at recognizing solutions to problems when we encounter them. Surely some of us are also good at inferring a solution by analogies; when being presented with an analogical version of a solution, some can abstract the principle and apply it more generally; but this is not the common case, as was shown by Maier.

6.3 The Principle of Autonomy
This far I have been using the term autonomy without a proper definition. As it is the essence of this thesis, I will take some space to discuss it here.
Based on reactions from Computer Science students, I have started to suspect that the word autonomy can give misleading associations, a suspicion that is shared by others (Stensson and Jansson 2014). It is probably not uncommon for a computer scientist to perceive a robot that is making its own decisions, based on some algorithm, as autonomous. This attribution falsely implies that autonomy has something to do with automation, which is close to the opposite of the actual meaning of the term.

Often autonomy is treated as a value, and is as such peculiarly synonymous with freedom; it is a condition that we should enjoy, which should be protected or given to people and other entities. However, autonomy is not meaningful unless the entity that enjoys it also has the ability and the possibility to reason autonomously. Autonomy has cognitive and structural preconditions.

In contrast to the mental processes that cognitive psychologists investigate, many philosophers differ between two possible mental dispositions. The first is the instinctive, habitual and authority bound attitude that leads us to believe that we are not really making decisions. The second is a deliberate, purpose-seeking attitude, which requires systematic reflection about advantages and disadvantages in different courses of actions. The former is called heteronomy and the latter is what I mean by autonomy. I will argue that the autonomy of the individual is essentially her ability to reason as a free agent, regardless of any authority or inclination.

Within Human Factors Engineering, autonomy is considered important because it is claimed to lead to safer, more productive and healthier work. To this, I would like to remind about the philosophical position that autonomy is that which constitutes human dignity. A pragmatic justification is that autonomy is necessary for being able to imagine alternative courses of action, i.e., to be creative.

I anchor my interpretation of these notions in the philosophy of Immanuel Kant (1785/2004) and the psychological inquiries of Jean Piaget (1932). There is communicative power in presenting concepts as dichotomies and we can observe endless examples of such in science, but we should not accept them blindly. Reality is usually more complex than the models we create to represent it. Nevertheless, some models are useful.

For a long time, I did not think it was necessary to make a distinction between, on the one hand, heteronomy and autonomy, and, on the other hand, the System I and System II reasoning processes that represent a more well-known cognitive model of human thought. That was up until I read Thinking Fast and Slow by Kahneman (2012) and realized that the System I and System II thinking processes do not account for all aspects of decision making, in particular not those that regard social aspects and relations to our surroundings. I will here make clear how these two taxonomies for human problem solving relate and how they differ. Figure 6.3 presents autonomy as a function of System II reasoning and a purpose-seeking independence of rules.
While reading the following, bear in mind that for neither the spectrum between heteronomy and autonomy, nor between System I and System II processes, do I regard the processes as mutually excluding. Heteronomy and System I thinking represent the usual way of thinking, which will be active, also when we reflect. I am not particularly interested in an exact model of brain functioning, but I find these constructs useful to create awareness of how to design artifacts and work processes so that we can make better decisions. The map does not change the landscape but can help us navigate in it.

6.3.1 Definition

The definition of autonomy that I am presenting here differs a bit from a colloquial use of the term, which is to denote entities that are self-governing. Neither do I see autonomy as just a value among others, which seems to be suggested in the Value-Sensitive Design (B. Friedman and Kahn Jr. 2003; B. Friedman, Kahn Jr., and Borning 2008) approach. Self-Determination Theory (SDT) offers one way to understand autonomy, that is a bit more elaborate:

[...] a person is autonomous when his or her behavior is experienced as willingly enacted and when he or she fully endorses the actions in which he or she is engaged and/or the values expressed by them. People are therefore most autonomous when they act in accord with their authentic interests or integrated values and desires (Deci & Ryan, 1985, 2000; Ryan, 1995). deCharms (1968) described a person who acts autonomously as an origin of behavior, because, when autonomous, a person feels initiative and stands behind what he or she does.” (Chirkov et al. 2003, p. 98)

The aspect of this definition that I want to point out is that it regards autonomy from the individual’s point of view, rather than from an observer’s. Although promising, there is one detail that makes it differ from how Kant defined the term. In SDT, following desires are included in the characteristics of acting
autonomously, whereas Kant explicitly states that desires occlude the clarity of reason and thus reduces the possibility to act autonomously. This may seem like a petty detail but it is actually very important since desires that conflict with autonomy can be induced from the outside, by, e.g., advertising, but also from inside, by, e.g., addiction. I will return to this.

For the sociologist Zygmunt Bauman, autonomy seems to be best understood as something of a social position in which one can find oneself. To be autonomous means to be an end and not a mean for someone else’s end (Bauman 1992). For instance, if your boss decides how you may act, you are not autonomous. This is nuanced with a perspective also on internalized authorities. If your habits have become so ingrained that you cannot reconsider them even when they make you miserable, you are not autonomous. Consequently, also Bauman suggests that autonomy requires awareness of the own interests.

Following the argumentation of Kant, I regard autonomy as an unattainable ideal mental disposition; as a noumenon; as something that we cannot fully circumscribe with our limited reasoning abilities. We can approach it but we simply cannot conceive it in its pure form. Our possibility to perceive and manipulate the world is constrained by our physical abilities, but perhaps more importantly, cognitive and dispositional factors constrain our ability to think in an unbiased way.

In the case of cognitive factors, it is useful to think of the two different cognitive processes that have been introduced above: System I: the instinctive, automatic, and effortless reactions to stimuli; and System II: the calculative, logical, and effortful computational process that we employ when we think hard about something. This second process is crucial for autonomy: if we cannot employ System II thinking, we stand no chance to make reasoned decisions.

In the case of dispositional factors, a dimension emerges in the attitude toward norms and rules. Jean Piaget wrote: “All morality consists in a system of rules, and the essence of all morality is to be sought for in the respect which the individual acquires for these rules” (Piaget 1932, p. 1). On the one extreme end, we place complete obedience to any command that is imposed on the individual, whether this derives from external or internalized rules. On the other end, we have complete independence from such influence. Piaget used this spectrum to understand the moral development of children: “[...] the main point is to find out whether one may legitimately alter rules and whether a rule

4. "The question then, 'How a categorical imperative is possible,' can be answered to this extent, that we can assign the only hypothesis on which it is possible, namely, the idea of freedom; and we can also discern the necessity of this hypothesis, and this is sufficient for the practical exercise of reason, that is, for the conviction of the validity of this imperative, and hence of the moral law; but how this hypothesis itself is possible can never be discerned by any human reason. On the hypothesis, however, that the will of an intelligence is free, its autonomy, as the essential formal condition of its determination, is a necessary consequence." (Kant 1785/2004, emphasis added)
is fair or just because it conforms to general usage (even newly introduced), or because it is endowed with an intrinsic and eternal value” (Piaget 1932, p. 15).

6.3.2 Jean Piaget and The Rules

In order to understand how autonomous thinking is developed in humans, the work that Piaget (1932) did on children provides a good starting point. Later psychological research on moral development in adults build largely on his theories. The main contribution to my work is a functional definition of autonomy as moral maturity; as an insight about what rules are, that they are created by fallible humans, and therefore possible to alter and improve if necessary.

For children, the learning curve is fairly steep and for a patient observer the development is noticeable. The cases studied by Piaget in *The Moral Development of the Child* (Piaget 1932), regarded simple games played by children, where the “spirit of the game” was the very reason for setting up rules. This was, nevertheless, not apparent to the children until rather late in their development.

Piaget observed four developmental stages in how children apply rules. During the first stage, rules are not yet cognized and it is a bit constructed to talk about deliberate application of them. At this first stage players are more or less unaware of their co-players and the aim is mostly to enjoy the physical stimulation from playing, the sensation of marbles and the fascination of experiencing physical laws at work. Communication could be described as “collective monologues”. This type of communication remains into the next stage, which is characterized as egocentric, since the child is neither interested in, nor aware of, how the play affects or connects to other people. Rules are learned from outside and are considered as eternal, and the child imitates the play only for its private satisfaction without regard to the social part of playing.

Later on this develops into the third stage where an incipient cooperation is arising. Rules embody a social contract and negotiations can be accepted if they serve the purpose of unifying differing rule sets. Each player tries, within the boundaries of the rules, to win the game. Finally, as the fourth stage is entered, it becomes apparent to the child that rules are codified and part of a social structure. They are understood as a property that is shared within the community and that the application of them is a way to achieve maximum enjoyability from the game. There is consensus, on a fairness basis, about what alterations to the rules are accepted and under which circumstances this can occur.

Along with the development of rule application is the development of consciousness about rules. This is “even more elusive in detail” (Piaget 1932, p. 17), as Piaget puts it, but three approximate stages can be identified. At first,
rules are considered at most as interesting examples of something that does not really apply to the individual. The children are at a stage which could be categorized as pre-social, in the sense that they are yet neither aware of, nor interested in the value of cooperation. During the second stage, rules are acknowledged and treated as divine and untouchable. They were once made by adults and will last forever, and the child is not comfortable with any alterations to them.

Finally, during the third stage, the rules are revealed as purposeful means of interaction. The children are aware of the fact that rules can exist only if every player accepts them. If a rule is generally considered as unfair, alterations will be accepted, given that mutual consent can be reached.

What is important in these two parallel developments is the correlation between them and how their interrelation affects the development of awareness. The argument for this insight is, however, not very apparent. Piaget sketches how the two developments interact; by application of rules pushing forward the consciousness of these, and the consciousness creating a foundation for the next stage in application, and how imitation of patterns works for adopting and eventually mastering the rules.

[...] the relations which exist between the practice and the consciousness of rules are those which will best enable us to define the psychological nature of moral realities (Piaget 1932, p. 3).

The mental process (consciousness of rules) of digesting the conditions that are already practiced (application of rules) lags behind and therefore the child will display a difference in attitude toward rule violations and alterations, depending on how it perceives the alterations. As an example we can take a child who is at stage two of the application of rules. Before the child becomes conscious about the origin and consistency of rules, alterations are accepted, but the reason is not the same as in the third, cooperative, stage; the alterations appear equally divine and normative as the rule itself. Acceptance is not based on an understanding of the purpose of the rule, as in the later stage.

This development is usually interpreted as occurring during the youth era, but this is an unnecessary delimitation. The theory becomes more general and helpful, if the stages of application are understood as mental functions—of increasing complexity—within specific domains (Kavathatzopoulos and Rigas 1998; Kavathatzopoulos and Rigas 2006). This view is supported by Lourenço and Machado, who concludes about critics that

[...] they often assume that Piaget considered age a criterion of developmental level, whereas for Piaget the key element was the sequence, not the age, of cognitive transformations […]. (Lourenço and Machado 1996, p. 147)

The reason for the common interpretation of timely stages is mostly pure convenience. It is hands-on and specific, and has a historical application within
child psychology where the stage model makes features more easily distinguishable. However useful, this misinterpretation has led to a lot of criticism of Piaget’s theory. It is clear that the development of morality applies also to adults, but it is, however, not as easily observable as in children. It is difficult to isolate a specific stage of development for a certain task, and the discretization into stages is obscuring the sound part of the theory. Within different fields, adults have developed their maturity differently, but the progression still follows roughly the pattern observed in children when we are allowed to adopt a purpose-seeking disposition.

An example from fiction might serve to illustrate this line of reasoning. In Orwell’s 1984, Big Brother systematically altered historical facts. News reports, novels, song lyrics, paintings, etc., were continuously revised to give the impression of Big Brother always being consistent and always making correct predictions. One interpretation, offered to the protagonist, Winston, as a false lead, is that this was done to give the impression of infallibility. “For to change one’s mind, or even one’s policy, is a confession of weakness” (Orwell (Eric Blair) 1949, p. 148).

Another, more psychological, interpretation is that it was done in order to prevent the citizens from reaching Piaget’s third stage of rule consciousness; that of autonomy. By dogmatically asserting truths, and by creating uncertainty about the own memory of facts, the citizens are not able to form a coherent mental model of what is happening. The only verifiable accounts of historical events are those which are given by Big Brother. In order to maintain the trust in the authority, these accounts do not have to be true, in the sense that a claim corresponds to an actual phenomena—it is in fact better if they are not. A person who does not trust her own reasoning is not able to object. The continuous manipulation of history is, thus, serving the motive of sustaining the citizen’s subordinate attitude toward the authority.

The mental disposition that we are interested in here, is that of moral maturity. By observing the attitude that children of different ages have toward the rules of a popular marble game, Piaget arrived at conclusions about how morality is learned. Premises for this are

1. that ethical systems essentially are rules, and the attitude toward such rules constitutes the essence of morality; and
2. that the rules of the marble game exist to make it more enjoyable to play.

These two premises lead to the conclusion that the children’s attitude toward the marble game is analogous to moral maturity, especially as the children eventually learn the true purpose of them and how to appropriate these.

We are in the presence of rules which have been elaborated by the children alone. It is of no moment whether these games strike us as ‘moral’ or not in their contents (Piaget 1932, p. 2).
What was significant for Piaget, was the relation between children’s actions and their consciousness of those actions. From infancy to adolescence, the attitude develops. The first step is learning, memorizing, and obeying the rules mechanically. Then, for a while, these are considered to be divine and eternally valid, and finally the children arrive at a point when they realize that the purpose of the rules is to make the game enjoyable and, consequently, should also be malleable. This final stage: to realize the malleability of rules, is reached only by reflecting on their purpose. This is the same mental disposition toward technology—as reified rules—that is encouraged by, e.g., Lawrence Lessig:

[...] all of us must learn at least enough to see that technology is plastic. It can be remade to do things differently. And that if there is a mistake that we who know too little about technology should make, it is the mistake of imagining technology to be too plastic, rather than not plastic enough (Lessig 2009, p. 32).

Kohlberg (1985) brought the theories of Piaget to practice, and took part in funding schools where students created and managed their own rules through democratic processes. In these schools, Kohlberg observed that when the rules were created by the students, they were also respected. By gaining insight into the creation of rules, the students developed their moral reasoning abilities. There is, however, a fundamental difference between Piaget’s and Kohlberg’s approaches. While Piaget studied attitudes toward rules, in order to determine how children construct an awareness of their function, Kohlberg worked from an assumption of what a morally developed individual is. For Piaget the mental disposition, revealed by arguments given for certain conduct, was the important aspect, whereas Kohlberg regarded also the conclusions as significant. This has been rightfully questioned, as founded on a biased set of values (Gilligan 1982). Nevertheless, it functioned to show that the theory is generalizable also to older.

Kavathatzopoulos and Rigas (1998; 2006) have later shown that the theory applies also to adults: people who hold high positions in organizations, who can be assumed to be well trained in decision making under uncertainty, show a more holistic and investigative attitude toward moral rules than the reference group consisting of students, who showed more dogmatism and authority obedience.

The generalization of the theory is that moral rules serve the purpose of creating order in communities, and that an autonomous, purpose-seeking attitude replaces conformity as we mature morally. Most of us are, however, not aware of this development process, and still at adult age we tend to refrain from questioning the universality of moral rules. The purpose of moral rules is not explicitly given, so it is difficult to acquire an reflective attitude toward them. It requires taking a philosophical position.
Nevertheless, moral maturity, i.e., autonomy, is very useful in situations where we cannot readily identify any precedents, and have a hard time to determine what is right to think and right to do. This is crucial for decision makers on any level. The self-doubt that comes along with the responsibility for autonomously reached conclusions is both an inevitable symptom of autonomous reasoning, and a strong motive to further engage this ability.

6.3.3 Autonomy and Functional Fixedness
Another way to understand autonomy, is as freedom from functional fixedness; as the ability to see the properties of a situation instead of specified functions. Rules are typically functions that serve some purpose, which may be unknown for the one who is subjected to them. Autonomous thinking means that new mental functions can be created to fulfill a purpose that is explicit, based on the properties of a situation—including the possibilities and the risks in certain courses of action. Here, however, we often have to fumble in the dark: what is the purpose of existing rules, procedures, and technology? Essentially, this boils down to the fundamental question: what is the purpose of being?

In this thesis I postpone the latter question by focusing on how to create necessary, although not sufficient, preconditions for identifying answers to the former: in order to be able to identify a purpose; and to appropriate rules, procedures, and technology so that they better serve it; an individual must be allowed to think autonomously. In order to develop the ability to think autonomously, the individual must be allowed to exercise her autonomy.

From an academic point of view, it may seem rather straightforward to stimulate the ability to reason autonomously. Fundamentally, it is about countering cognitive biases and supporting the construction of a mental model that is sufficiently correct for making informed decisions. In practice, however, it is not as easy. Our thinking is constrained by many factors on different levels: data may be presented in a biased way; the operations that we are allowed to perform may be limited, and the routines that we follow may reduce our possibilities to form enough understanding to be in control over our decision making. The key problem is, however, that we quickly become accustomed to phenomena and practices, which we subsequently accept without deeper reflection. Things “are” in a certain way, and most of the time this is very useful, as it relieves us from unnecessary cognitive processing. The problem is that it is difficult to “unknow” what we know, even when that is not suitable (Stroop 1935). Familiarity and already established concepts are a double edged sword.

6.3.4 Consequences of a Strict Definition
One consequence of my stricter definition of autonomy is that no existing machines can be regarded as autonomous, even if they make decisions inde-
pendently of a human operator. They are merely calculative in an automatic sense, as they, without reflection and without purpose, follow some set of rules that are created for them. This claim may, however, seem a bit hollow as there actually have been constructed machines that to some extent are capable of formulating rules by themselves. Recently a breakthrough has been heralded in news across the world, about a machine that, after training without human supervision, has learned to detect if an image contains a cat face. What this sensational result translates to, for an academic, is that from a database of 16 million images, belonging to 22,000 different categories, the machine managed to make the correct categorization for 15.8% of the images (chance would give about 0.045%) (Le et al. 2012). This is, indeed, remarkable, but is yet not close to any consciousness about rules.

Even if a machine eventually will be able to classify images with full accuracy, this task is qualitatively different from using such knowledge for some higher-level purpose. As an illustration, consider compiling a dictionary out of a set of unordered words. This task is trivial for a computer, and without supervision it would likely come up with a very large number of different ways to arrange the words: according to word length, Levenshtein distance, occurrence frequency in text, etc., many of which would make little sense to us; the alphabetical ordering from left to right that we are accustomed to, is just one of the possible ways. The reason why we have chosen this convention, which also presupposes an agreement over letter order, is that it was once useful for looking up words in physical dictionaries. Today it is less necessary but since we are accustomed to it, we make use of this convention in various situations, like in staff ledgers. Nevertheless, it is not the ordering itself that is interesting but what we can use it for; what it enables us to do.

Classification, or categorization, is, without doubt, an important part of human decision making, and it helps us make sense of the world. One of the assumptions that I have made in this thesis is that classification is not a strictly cognitive, but also a social process; that much of it results from a process of appropriating language, based on functionality of artifacts and concepts, with usefulness in mind. Furthermore, language is slowly evolving to match changing preconditions. Words acquire connotations that determine their appropriateness in different contexts, and classification follows suit. This is a process that requires some form of calculative ability but also a purpose; a direction. If a machine were to develop a similar agenda, it is not given that it would be the same as ours.

But if technology can be shown to be nonheteronomous, what does this say about human will? Ellul is explicit on this point: “There can be no human autonomy in the face of technical autonomy.” In his eyes there is a one-for-one exchange (Winner 1978).

For me, personally, the purpose may seem rather reactionary: to preserve a conception of humans as ends in themselves. While we, as a species, have
developed the notion of rationality and use it to prescribe ideal human behavior, we also need to question what really is ideal. The last hundreds of years, rationality has been associated with logic, efficiency and disconnection from emotions. Psychological research, especially within neuroscience, has lately started to question this ambition. Definitions of rationality should have ecological validity, i.e. be based on what is successful in a system, rather than just in an experiment, and it seems like emotions play an important part in successful decision-making. After surveying literature that indicates a positive correlation between brain malfunctioning and rationality, Hertwig and Volz conclude that “some recent studies on the link between abnormality and rationality do invite one to challenge the very sanity of some classic norms of rationality” (Hertwig and Volz 2013, p. 548). This undermines System II thinking as the ideal for rationality and leads to another consequence of the definition: neither humans who are blindly following rules can be considered as autonomous. This may seem like a harsh statement but the difference between a machine following rules, and a human following rules is that the human, indeed, has the capability to decide differently. Autonomy is thus not a property of an entity that is always present but more like a potential that can be used or not. We can, and do, voluntarily trade in our autonomy for, e.g., efficiency, security, and perhaps most of all: to avoid responsibility for our decisions. This is often both necessary and good, and we should not adopt an attitude where we are constantly questioning mundane tasks, like how to best prepare breakfast. However, when we need to act autonomously, we should be ready to do so, and we should not be hindered by technology. Ideally, technology would prepare us to be in control over our lives and our activities. What is interesting for designers, is, thus, to understand when it is necessary to allow us to make use of our potential and how to accomplish this.

But that the principle of autonomy in question is the sole principle of morals can be readily shown by mere analysis of the conceptions of morality. For by this analysis we find that its principle must be a categorical imperative and that what this commands is neither more nor less than this very autonomy (Kant 1785/2004).

6.3.5 Heteronomy and the Normative Matrix

According to SDT, the opposite of autonomy is not dependence but rather heteronomy, in which one’s actions are experienced as controlled by forces that are phenomenally alien to the self or that compel one to behave in specific ways regardless of one’s values or interests. (Chirkov et al. 2003, p. 98)

5Keith Hammond would probably not agree with using the term ecological validity for this purpose, by insisting that Brunswik intended this to only regard cues to information that an entity uses to cope in an environment.
For many human endeavors there exists some agreed-upon way to accomplish it in an ideal way. I think of this as a Normative Matrix. This includes everything that we consider to be normal: vanilla ice cream in the summer, popcorn in the movie theater, Spaghetti Bolognese, day jobs, etc. The Normative Matrix does not force us to reflect—within its domain everything is normal and we can act effortlessly and feel confident about it, which is essentially good. However, the Normative Matrix can also be a restraint. It may make us more reluctant to question norms and truths than we ought to be and it may prevent us from making the right thing.

When we make decisions without being aware of there being a choice involved, we are acting heteronomously in its most straightforward meaning. To this, I would add situations in which we are aware of choices but nevertheless for some reason refrain from making a deliberated decision. This could be due to social pressure, overconfidence, or other cognitive biases offered by the Normative Matrix. Most of the time this is good and I will use an example to illustrate it.

Imagine a woman driving her car on the expressway at two o’clock in the morning. She is on her way home from visiting her parents in Hamburg and has less than an hour’s drive left before she can finally sleep in her own bed. She has been driving since the morning, with only short breaks to eat and visit the restroom. Traffic is low on the highway, and only occasionally does she see other late night drivers. There is some distance between the opposing lanes so she does not have to care much about blinding approaching cars with her full beam and the car almost automatically follows the road. Her GPS navigator is informing her about which way to drive so she can comfortably sit back and enjoy the minimalist techno music that is streaming out from the car speakers. All of a sudden, she is shaken out of her comfort. Just by chance was she glancing at the passing road signs and noticed that she had missed the exit leading back to her hometown Uppsala. She is now heading towards Västerås, which from her current position is in the completely opposite direction. She knew that her GPS navigator’s map had not been updated for some years but this is the first time that the instructions actually fail. On the map in the navigator, it now appears as if she is not driving on the road anymore. After a deep sigh, she realizes that she has to keep driving until an exit lets her drive back and take off toward Uppsala.

This is the conclusion that most drivers would arrive at but there are, indeed, other options available. One could brake and back up to the exit or one could make a u-turn in the middle of the empty road to hastily drive back to the exit. Such solutions, for good reasons, seem very risky. Highways are a typical setting where people conform and expect others to conform. Attention for deviances is probably low, as drivers, after many years of driving, have become used to not waste cognitive effort on paying attention to other cars as the main potential threat.
A negative way to see heteronomy is that it is fixated associations. It means that the thinking is constrained by previous knowledge and authorities, rules, biases and heuristics. It can be considered as a form of learned helplessness, with the difference that it is not actions that are constrained but our reasoning. However, in most situations heteronomy leads to perfect conclusions. It is a perfect, deterministic match between a predetermined purpose and decisions. In such situations, we should not force ourselves to do unnecessary reflection. On the other hand, when we are dealing with non-prototypical situations—situations in which we cannot readily apply experiential knowledge—the mental shortcuts can misfire and create erring decisions. One great challenge is to realize when we find ourselves in such a situation, and to have the confidence and training to handle it satisfactorily.

6.3.6 Autonomy and Understanding

One important precondition for being able to act autonomously is a well-developed mental model of the mechanisms in a situation. Without understanding, an individual is not able to make deliberate decisions. This point can be introduced with a somewhat lengthy quote from Donald Norman:

I want a system that is enjoyable to use.

This is an important, dominating design philosophy, easier to say than to do. It implies developing systems that provide a strong sense of understanding and control. This means tools that reveal their underlying conceptual model and allow for interaction, tools that emphasize comfort, ease, and pleasure of use: for what Illich (1973) has called convivial fools. A major factor in this debate is the feeling of control that the user has over the operations that are being performed. A “powerful,” “intelligent” system can lead to the well documented problems of “overautomation,” causing the user to be a passive observer of operations, no longer in control of either what operations take place, or of how they are done. On the other hand, systems that are not sufficiently powerful or intelligent can leave too large a gap in the mappings from intention to action execution and from system state to psychological interpretation. The result is that operation and interpretation are complex and difficult, and the user again feels out of control, distanced from the system. (Norman 1986, p. 49)

Norman uses the words *understanding* and *control* instead of autonomy. These concepts are, indeed, important prerequisites for being able to think and act autonomously, but they are defined in relation to technology, not as being intrinsic to the individual. They are not explicitly meant to address a purpose that the user may define, and are, thus, not explicitly allowing autonomy. For instance, also slot machines are designed to give the user an impression of understanding and control. If only the function of the machines was important, they would be black boxes that deduct a random percentage from your account’s balance. The perception of understanding and control is, thus, not
enough. Questions like *why does the user need control?* are not satisfactorily answered without a philosophy that takes the intrinsic values of individuals in account.

### 6.4 Arguments against Autonomy

The autonomy of an individual needs to be balanced with allowing others to be autonomous. This is essentially the purpose of ethics. An illustration of when this is neglected is a manager who wants to take responsibility for every aspect of her enterprise. To accomplish this, she may use ICT systems that allows her to make sure that operations are running as expected. However, if these systems are controlling her employees too strictly, it may violate their autonomy. They will be treated as means for her end, which, according to the present theory, may reduce their ability to reason and take responsibility.

Autonomy is also expensive—cognitively speaking—and uncertain. It does not produce the same predictable outcome as heteronomy. This unpredictability is stressful, and there are indications that if expectations on acting autonomously are not balanced with social support, it may lead to stress-related diseases (Johnson and Hall 1988):

> What was not predicted by the earlier demand-control model, however, was the elevated PRs found among active-isolated workers. Although we can only speculate as to the explanation for this finding, it is possible that in some working situations, high levels of control may accentuate rather than reduce the impact of demands. Our indicator of control may actually be measuring responsibility pressure, which in some occupations might constitute another component of job demands. This suggests the need to further refine our measurement of work control in order to at least distinguish control as a resource from responsibility as a demand (Johnson and Hall 1988, p. 1341).
7. The Human Context

7.1 Cognitive Processes versus Rituals

In this thesis I am focusing on processes. A dictionary definition of this term is “a series of actions that produce something or that lead to a particular result” (dictionary 2014). This is true for manufacturing processes, as an important quality of these is that they produce consistent results. When it comes to cognitive work processes, it is not obvious that it is the same. I realize that it is difficult to argue for a definition in which the activities are evaluated independently from the result that they lead to, although this is exactly what I would like to accomplish: I see a cognitive work process as a theoretically justified series of actions to address some question, which may be aiming at producing a desirable result, but more generally, the result is a consequence of the process. It is, thus, not a particular result that is interesting but the way that the result is reached.

An example of this is calculus. Given some quantitative data, a mathematician starts processing these. She may have a determined goal, like calculating the area of a rectangle, but she is not interested in producing, for instance, $5m^2$ as a result; she is interested in finding out what the area really is.

Another example is the writing of this thesis. The goal is determined: I need to produce a long and novel sequence of words, but it is not determined in advance what this sequence is. My purpose affects it—there may be a message that I want to get across, I may want to show that I learned the trade, or I may just want to earn a PhD—but it is also affected by events that happen during the writing process—what I read, what I realize, and so on. Different purposes require different processes so the process must be chosen to correspond to the purpose. Otherwise, the process is merely a ritual that is to be performed.

If we now turn to rituals, we can observe that they are context-dependent—often habitual—ways of doing things. The difference between cognitive work processes and rituals is to my understanding only in their actual effect. Ritual actions are symbolic, and are often performed without getting observable feedback; it is the experience of performing them that matters. A person who saves a text document several times before exiting an application is performing a ritual that increases her feelings of comfort. Ritualized behavior is a symptom of lack of understanding (Gigerenzer, Krauss, and Vitouch 2004; Dulaney and Fiske 1994).
7.2 Expertise and Being Human in Processes

Processes should not take the form of rituals. Very detailed processes easily become subject to this fallacy. Gilbert Cockton has argued that there is an overly myopic focus on process rigidity amongst HCI researchers (Cockton 2013). What is considered important is to perform chosen methods correctly, regardless of how much information they yield and regardless of how successful they are in practice (Hornbæk 2010). This is not necessarily always wrong, as it makes processes predictable and enables generalization. One of the stronger arguments for focusing on getting the rituals right is that a lot of assumptions and reasoning have gone into the details, which the user of the method does not need to understand. Nevertheless, accentuating detailed method protocols may produce unwanted effects, especially on the cognitive abilities of designers. If prospective experts are expected to always follow detailed instructions, they will not train their abilities to judge and reason, and they may not be able to take responsibility for the outcome, as this is dissociated from the work that they do.

Ethnologically oriented HCI scholars like Paul Dourish have pointed out that the way that ethnography often is used in HCI misses the point of doing it in the first place (Dourish 2006). Producing an ethnography is an end in itself and using ethnographic methods for informing design is misguided, as it is then directed by a particular focus instead of being an interpretive reportage about what is going on and what is experienced in a situation. Posing expectations on a result will affect what the ethnographer is observing and what he or she will report. This is not to say that there is no value for designers in ethnography. On the contrary, a rich and nuanced interpretation of an ecosystem is invaluable if your mission is to affect it, but this understanding cannot be digested into a few concrete implications. And if someone who does not understand these assumptions follows the same protocol, it becomes a ritual; a cargo cult.

As an example, consider a person writing in Chinese without knowing what the symbols represent or what are the significant information-laden features of the strokes. This means that if the person does not follow a prescribed ritual slavishly, the produced squiggles may not be intelligible to a reader. Any deviation, be it for artistic reasons or due to neglect, risk obfuscating the message that was to be conveyed. Now, even though this type of reasoning makes sense and seems to support the idea of meticulous attention to getting the procedure right, it is still a peculiar stance. Why would we ever want someone to do something that he or she does not comprehend?

Admittedly the outcome is impossible to assess and predict in the moment of development, but what is crucial for me is that a method should not prevent the user of it to engage her critical thinking. It may seem paradoxical if I now claim that my supervisor Iordanis Kavathatzopoulos is of the position

---

1This is essentially the Chinese room scenario, described by Searle (1980).
that when it comes to moral decisions, these should not be assessed by their outcome, but solely by the method that was used to reach them (Kavathatzopoulos 2013). However, his approach is Kantian, and the method is exactly that of autonomous, critical reflection. The purpose of my work is to encourage this way of thinking in design practice.

Fundamentally: the more carefully a procedure is detailed, the less it requires users of it to engage their critical, autonomous thinking. Cockton is observing that designers are treated as if they were just executing a set of instructions, much like the thinking was in the era of cognitive engineering (Cockton 2013). Judging from the attempts to introduce detailed methods into design practice, it seems like the main problem in design is the designers’ inability. Why else would they be expected to follow rules so rigidly?

Although I agree with Cockton’s problem description, I do not believe that his remedy—to introduce richer and more inspiring vocabularies—is going to solve the problem. To use a more diverse language when describing design activities, depending on who we are addressing, may stimulate the thinking and understanding of stakeholders who do not identify as designers, but it may also make the process less distinct and more prone to unsubstantiated opinions. The point should be to allow designers to be more autonomous, so that they can develop expertise, but the point with establishing a fixed terminology is, after all, to encourage a shared repertoire of background knowledge, which is also part of that.

### 7.3 Morality

One of the first questions that I was struggling with, when I was doing my Master thesis, was whether it is possible to teach people to be more moral. At the time I thought that morality is something that people have to a certain degree (where I assumed that more is better) and that this is rather fixed for individuals. My reasoning at the time suggested, for instance, that a person who grows up as a bully will continue to be a bully in spirit through adulthood. She will just become better at hiding this, since it is usually frowned upon in public. This is probably true for many people but I did learn something from my struggles and found that philosophy is a great tool to curb the influences that a bad upbringing may have caused.

If I venture to assume anything about people, it is that we are different. That we are all individuals is a rather optimistic claim but at least something of the kind I believe is true. Some hurt other people on purpose, most do not. Yet others act in incredibly self-sacrificing ways.

It would be tempting to conclude that this has all to do with how moral the person in question is. Philosophers have thought about this for some time now. Aristotle (1910), who is something of an icon for virtue ethics, claimed that a person, by proper exercise, could become more virtuous if she wants
Virtues are what constitutes good character and are, according to Aristotle, generally found in the moderation of opposing vices. For instance, courage is by this logic to be found between cowardice and rashness, and pride between vanity and undue humility. Deficits in character can be the result of three different conditions: ignorance, incontinence (lack of self-control), or vices. The first one is easy to remedy by education, the second one is a condition that can be excused whereas the third one is voluntary, and thus a sign of truly bad character.

What can be said with relative certainty is that these virtues and vices reflect what was considered good and bad in Greece around 350 BCE. They have not changed much since, but we may add some present-day vices to the blend. Between the paranoid and the carefree. Between the control freak and the day-by-day, between the micro manager and the laissez-faire leader we can find the virtuous individual.

Now, I will claim that virtues are auxiliary to what is truly moral. I will go even further to claim that morality is a construct to make sense of what is rational for a complex social system.

A system perspective is important for understanding morality. If we observe that the freedom—the possibility to decide and act without externally imposed limitations—of one entity in a system can be constrained by the decisions of another, and if we value freedom, there should in a system of entities exist mechanisms to determine what are permissible decisions. We can make some further deductions:

First, the mechanisms for balancing freedom of individual entities are only necessary if the entities can make independent decisions. In a deterministic system, in which the entities never have, or realize, options, such mechanisms are never employed. Neither are they useful in a system where the freedom of one entity cannot be constrained by the decisions of another.

Second, the mechanisms for balancing the freedom of individual entities are only relevant if the entities can, and do, base their decisions on these mechanisms, i.e., if the entities are not deciding deterministically, yet choose to regard the system in their decisions.

Third, from an evolutionary point of view, a system of free and independent entities would risk destroying itself unless there exist mechanisms for governing decisions in a way that preserves the system. It is thus likely that humankind has developed such mechanisms and that these are intrinsic to us. They are what we refer to as morality.

Consequently, morality is the respect for the system of people that is internalized in an individual. Immanuel Kant has a lot to say about that, but first the question of how morality is communicated and distributed in these systems.
7.3.1 Moralizing

Moralizing is the activity that corresponds to morality. I acknowledge that most people have a narrower definition of the term than what I am to propose. This is because I ascribe moral value to any choice situation and not only those that usually pop up when the word morality is mentioned. There does not have to be a dilemma for a choice to be morally relevant. This generalized view is purposeful for understanding how morality is transferred between humans. It is a process that we engage in joyfully throughout our short lives. Whenever we make a value statement about other people’s actions or choices we are participating: Nagging, agreeing, objecting, laughing, etc. are all different ways to express moral opinions. Also more subtle signaling—like smiling, ignoring, humming, changing the topic, not laughing, and looking away—communicate moral opinion. Moralizing is a natural behavior for a social animal inhabiting a social system. It is also a necessary behavior if the system is to be sustained. The entities in the system need to agree on a shared set of codes, norms and standards so that it is possible to predict the behavior of others and know how to act in order to achieve a desired result. Such rules are too plentiful to all be communicated in a concerted effort so they need to be realized in the relevant situation.² Moralizing is, thus, simply a way to communicate, negotiate and sustain a complex system of rules. For the reader who prefers to view things from perspective of a system, the implicit goal of a society (the system) could be seen as sufficient stability and predictability to guarantee further prosperity, at the same time as it needs to adapt to changing conditions. If there were no social conventions, the society would not provide security to its citizens, and would thus not be to any benefit for them, but if social rules were rigidly fixed, the society would not fulfill its purpose if conditions changed.

It seems like Immanuel Kant attributed the same important role to moralizing when he sketched his fundamentals of morality, although he used the expression “legislating member in the universal kingdom of ends” to describe what I usually call a moral agent. He realized that every rational being in a system has the power to influence the opinions of other beings:

> A rational being must always regard himself as giving laws either as member or as sovereign in a kingdom of ends which is rendered possible by the freedom of will. (Kant 1785/2004)

To prevent morality from following the wrong path (i.e. the path leading to a system that does not benefit its entities) he formulated the categorical imperative as basis for morality, the well-known formulation of which is: “So

²Besides, abstraction and written language is a rather modern invention if we regard the likely human history, so the enterprise to collect all rules was probably even more difficult for our illiterate ancestors. Thus a more convenient, less taxing method had to evolve.
act as if thy maxim were to serve likewise as the universal law (of all rational beings)” (Kant 1785/2004).

These two statements should not be interpreted in isolation. One may question whether we act as rational beings all the time but that does not reduce the implication that we are, by our everyday actions and decisions, (re)creating rules for our peers. My decisions affect yours and vice versa. Every member of a society takes part in forming the rules of the society and should realize their responsibility in this.

The practical necessity of acting on this principle, i.e., duty, does not rest at all on feelings, impulses, or inclinations, but solely on the relation of rational beings to one another, a relation in which the will of a rational being must always be regarded as legislative, since otherwise it could not be conceived as an end in itself. (Kant 1785/2004)

In our current society the role of moral agent is, however, not confined to individuals that interact directly with each other but includes all sorts of institutions and unidirectional channels: governmental as well as private. The Swedish philosopher Claes Gustafsson and the Canadian media scholar Marshall McLuhan make strong cases for that media is one of the most efficient moral agents (Gustafsson 1997; McLuhan and Fiore 2001). The former equates the effect of media with a megaphone: a (mostly) unidirectional channel through which some moral agent can impose opinions about right and wrong on the public (a description that could also be applied to much of social media). He is troubled with the oversimplification and exaggeration that results from several different megaphones competing for attention. Also McLuhan voices a similar concern:

…media, by altering the environment, evoke in us their unique ratios of sense perceptions. the extension of any one sense alters the way we think and act—the way we perceive the world. Were the Great Blackout of 1965 to have continued for half a year, there would be no doubt how electric technology shapes, works over, alters—massages—every instant of our lives (McLuhan and Fiore 2001, p. 147).

What is important to realize is that we are constantly involved in exchanging moral imperatives with others and that most of us are receiving more moral opinions than we are giving.

7.3.2 Moral Judgment

Moral situations are usually charged with emotions. Instinctively we feel repulsed or are provoked by behavior that runs counter to our idea of moral conduct. After deep deliberation and discussions (perhaps even involving a heated
debate) we sometimes eventually have to accept that, although offensive to us, some behavior and some decisions still must be considered as legitimate and sound. In cases where we do not have an urgent personal interest, such a time-consuming acclimatization process is acceptable and realistic, but when we do have vested interest in the conclusion, and have to reach it in reasonable time, we probably need help to structure our thinking. We need to deploy methods that help us counter our natural biases and subjective preferences so that we can assess the situation as objectively as possible. Indeed this is in the interest of any decision maker. Simply aiming for personal gain will not be considered as a sign of good judgment if the decision is ever scrutinized, rather the contrary.

Kant may appear a bit controversial when he claims that actions that for a lay observer appear moral, like for instance following the law or being fair, are not necessarily the result of a moral disposition, even if they would be in accordance with it. His argument is that every decision of a human being is based not only in our reasoning, but also in our inclinations. Consequently, it is not possible to distinguish whether we did good because we acted out of a moral disposition or out of self-interest.

Only those situations in which we act against our own self-interests can, thus, reveal a moral disposition. However, in practice such situations are impossible to find. Even if we, for moral reasons, would act against the norms of society and make ourselves appear despicable, we may still be acting out of the kind of self-love that is about feeling moral high-ground. Examples of this are religious practices that are frowned upon in a secular society. They may be in accordance with the scriptures of the religion, but that does not make them moral according to Kant’s definition. For him, morality was only to be determined by how decisions were made. Many have interpreted him as a dogmatist; as if he thinks that doing one’s duty is the highest goal. In a sense this is true, but the implications are not as one may think. The confusion about his position is to be found in misunderstanding of duty. Compare the two quotes below:

If now the maxims of rational beings are not by their own nature coincident with this objective principle, then the necessity of acting on it is called practical necessitation, i.e., duty. (Kant 1785/2004)

For duty is to be a practical, unconditional necessity of action; it must therefore hold for all rational beings (to whom an imperative can apply at all), and for this reason only be also a law for all human wills. (Kant 1785/2004)

The first quote can be taken to suggest that duty is imposed from the outside, whereas the second clarifies that duty is the foundation for what we want. For Kant, duty was not about blindly following the law or the orders of superiors but to act according to the principle of autonomy.
The familiar formulation of Kant’s categorical imperative—“Act only on that maxim whereby thou canst at the same time will that it should become a universal law”—has for many taken on the meaning: treat others as you would want to be treated by them. Many have also taken this to be a guideline to determine what is moral, and as such use it as a normative ethics. I believe that Kant instead had the ambition that I suggested above. In his writing he opposes dogmas and it would thus be peculiar if he were to suggest something similar himself. Rather than as a command, the imperative is categorical; unconditional in the sense that it is how we function as rational beings. I therefore understand *Fundamental Principles of the Metaphysic of Morals* to be a *description* of the necessary conditions to achieve a functioning system of agents that are not commanded from outside, but from within. What, then, is this command?

[… ] the moral imperative is expressed thus, that the maxims must be so chosen as if they were to serve as universal laws of nature. (Kant 1785/2004)

Morality consists then in the reference of all action to the legislation which alone can render a kingdom of ends possible. (Kant 1785/2004)

This legislation must be capable of existing in every rational being and of emanating from his will, so that the principle of this will is never to act on any maxim which could not without contradiction be also a universal law […] (Kant 1785/2004)

These quotations expresses the idea that for morality to be possible, individuals need to be able to internalize it—we must even want to do so. Being moral means that the will of individuals is in accordance with the purpose of a society, i.e., to sustain a kingdom of ends; a society of autonomously acting individuals. However, we do not always act in this way, as we are easily disturbed by other influences: feelings, impulses, and so on.

Reason then refers every maxim of the will, regarding it as legislating universally, to every other will and also to every action towards oneself; and this not on account of any other practical motive or any future advantage, but from the idea of the dignity of a rational being, obeying no law but that which he himself also gives. (Kant 1785/2004)

It was seen that man was bound to laws by duty, but it was not observed that the laws to which he is subject are only those of his own giving, though at the same time they are universal, and that he is only bound to act in conformity with his own will; a will, however, which is designed by nature to give universal laws. (Kant 1785/2004)

To conclude: Kant saw morality as that which meant respect for all individuals in a society, since only that can constitute a society that is worth to exist in.
This has become difficult in the current society, in which we can immediately identify many dominant influences that are in conflict with such a conception, both in private and public spheres.
Part III:
Theorizing about Artifacts

“[I]t must be possible for an empirical scientific system to be refuted by experience.” (Popper 2005, p. 18)
8. The Theorizing Process

In this part of the thesis, I will propose a protocol to aid systematic analysis and communication about how artifacts should be designed to correspond to the actual interests of users.

Even though I make use of the terms theorizing and theory, I do not suggest that the generated theories in themselves automatically become general design theories. I readily acknowledge such to be difficult to corroborate satisfactorily (Gaver 2012).

In 2008, Ben Shneiderman wrote an article for the Science Magazine, in which he calls for new methods for doing research—methods that take into consideration the uncontrollable nature of real life, in contrast to the controlled, unrepresentative environment of a laboratory.

Science 2.0 researchers who develop innovative theories, hypothesis testing based on case study research methods, and new predictive models are likely to lead the way. The quest for empirical validity will drive research beyond what laboratory-based controlled studies can provide, while replicability and generalizability will be achieved with greater effort through multiple case studies. (Shneiderman 2008)

The protocol that is proposed here is not specifically directed toward research, but it implies hypothesis testing and can be used for aggregating results of case studies.

Psychological studies have shown that the past experiences that people value the most are the ones when they felt self-esteem, autonomy, competence, and that they could relate with others (Sheldon et al. 2001). All of these values require autonomy, as described in Section 6.3. In order to feel self-esteem it is necessary that the individual has been able to decide responsibly. Competence can only be achieved through internalization of well-functioning practices, which, in turn, can only result from a proper understanding of the problems that are addressed. Relatedness requires being respected by other thinking, judging, voluntarily engaging individuals (Ryan and Deci 2000). The predictive model that Shneiderman is advertising for is thus in the present proposal to a large extent replaced by the principle of autonomy. If we have a clear conception about what autonomy implies, we can use that understanding to deduce what supports it and what hampers it and, consequently, we have the best possible tool to predict which designs will work and which will not.

However, the protocol does not explicitly prescribe designing for autonomy, so there are reasons for designers and developers (from here on DDs) to adopt
it even if autonomy is not accepted as the highest value. Recurrently, artifacts that fail to satisfy the intended users’ needs are produced, either because the artifacts are not well-designed enough or because they are solving the wrong problem. I can see at least three possible explanations for why the second problem occurs.

1. First, the DDs were not competent enough.
2. Second, the DDs were competent, but did not understand what the problem was that they were expected to solve.
3. Third, the DDs were both competent and understood what the problem was, but were, for other reasons, not able to produce the right solution.

The suggested protocol addresses the first two problems. The third suggests an organizational problem, which will be touched upon, but not explicitly addressed.

8.1 Refocus

Despite decades of research, literature output, and strong advocacy for usability, it is still common to encounter ICT systems that do not improve the situation for the intended users. In retrospective analyses, HCI scholars and practitioners will readily point at the lack of usability. Sometimes the criticism is extended: the systems were designed by the wrong methodology, or they were designed for the wrong tasks or goals, or by the wrong people. With all the existing knowledge, and all the established methods to safeguard usability, how come this problem recurs? It almost seems as if usability advocates have failed to impact practice.

The main reason, as I see it, is not the lack of good methods to get the user interface design right. The body of knowledge within HCI includes many accounts of methods that have worked well for certain applications. These and other methods are in widespread use among usability-aware practitioners, helping them to design artifacts that meet at least a minimum level of usability. However, there is also a common misconception (see Section 6.1) that it is possible to make an artifact usable by correcting its user interface. The strong advocacy for, and adoption of, usability testing methods is evidence of this (Greenberg and Buxton 2008).

Ironically, the present idea is a reaction to the tendency within HCI of proposing and embracing new methods to resolve issues in artifact design. What I have found lacking are explicit descriptions of what the premises are. For instance, what are the psychological and philosophical assumptions behind personas (Blomkvist 2003)?

I have this far in the thesis presented some thinking tools that have led me to make certain observations.
Much of research within HCI has been focusing on users, and how to construct systems so that these can create appropriate mental models of computerized artifacts. Some attention has also been given to organization of development: how different development processes, like the modern agile versions, address users’ needs (Larusdottir 2012; Paetsch, Eberlein, and Maurer 2003). The agile philosophy (Beck et al. 2001) is currently very popular within systems development, and it does, at least in theory, give a high degree of freedom under responsibility to DDs (Paetsch, Eberlein, and Maurer 2003). However, it is not explicit about the value of autonomy, so effects for users of the produced artifacts cannot be guaranteed. Furthermore, as I wrote in Section 6.1.2, it is not clear how DDs come to form mental models of the problems that they are to tackle. This is my target here.

What I propose is a lightweight participatory design protocol that is intended to help DDs develop their understanding of what is the right system to build. It is very simple to adopt, and is devised to counter the effect of cognitive biases. This is accomplished by refocusing attention, from the users to the DDs of an artifact.

8.2 What Is the Problem?

Neil Postman used to ask: “What is the problem to which this technology is a solution?”

The founding premise for the present idea is that when an artifact is created, be it an enterprise system or a single-purpose consumer product, the design of it is the DDs’ best possible answer to some question. This answer is naturally constrained by available resources (in the widest meaning of the word, including, but not limited to, money and time), environmental factors, and the knowledge and imagination of the DDs. To some, it might seem a bit odd to think of designs as answers, since it is not always perfectly clear what the question was.

On the other hand, if we realize a problem, it leads to concrete questions, and these questions can, in turn, lead to designs. This is what I wish to bring to the fore. Why design something that does not solve a problem? Or more generally, why answer a question that nobody has asked? If we, for instance, think about what the question was that lead to the typewriter, we may be able to solve that problem differently.

Such, and other related issues, I will try to make clearer. Another objective is to encourage both DDs and users to become more articulate about the purposes, values, and judgments that lead to the creation of an artifact, as these also affect design decisions.
8.3 Design as Applied Ethics

To design artifacts in an ethical way is important, especially if we consider design as an activity of changing the way people do things and, by that, of disrupting current practices. We need to take into account the “kingdom of ends”, i.e., people who deserve to be treated as autonomous, in the making of things. We need to assume that they have the ability to be rational, and give them the possibility to take responsibility for their decisions.

\[ \text{...the will is a faculty to choose that only which reason independent of inclination recognises as practically necessary, i.e., as good. (Kant 1785/2004)} \]

8.4 Conceptions of Good Design

Throughout this thesis I am developing an idea of what is good design. This, however, may differ slightly from the associations that the reader may have. The public conception of a well-designed artifact is a moving target. As possibilities increase, so do expectations. All of this happens in a complex interplay, in which technological advances almost enable something, which in the public mind then is interpreted as a promise for the future; as something that we may expect. We are, for instance, able to surf many webpages from our smartphones, which makes us consider those that are not accessible on the go to be badly designed. We can, with some effort, model and print rudimentary plastic things on 3D printers, which promises that we will soon be able to take snapshots of whatever we like, and produce exact replicas at home.

A similar tendency can be observed in design of computer systems that are intended to support work. Workers are naturally exposed to the wonders of ICT mediated entertainment and, consequently, they start questioning why their working tools are not as well-designed.

An obvious reaction is to regard the design of consumer technology as an ideal also for working tools. Surely, this has both merits (Franko and Tirrell 2012) and justification: consumer technology is presented to a market in which there is strong competition, so only those artifacts that have a competitive edge will survive the selection process. However, it is important to remember that consumer technology is not necessarily designed to empower the user, as a more likely motive is that it is designed to maximize the revenue of the company that offers it. In a free, competitive market, these incentives will sometimes be aligned, but that is dependent on consumers being informed, and that they are making rational purchase decisions. I guess it is safe to say that these premises are not always fulfilled. Furthermore, it can be difficult to identify what it really is that consumers love in their gadgets. This may be anything from functionality, the simplicity in handling, to the identity that is attributed to owners of a particular brand of artifacts. Finally, consumers’ requirements most likely differ from those of professional users. The consumer,
as a faceless abstraction, does not need to acquire expertise in the same way as professional users.

Gamification is another obvious response to the request (Deterding et al. 2011). The motto could sound something like: since games are so engaging, we could perhaps add some elements to work processes so that they resemble games, and that would increase the motivation for working. I am supportive of the ambition to make work more engaging, but I would like to see gamification addressing the underlying problems in poorly designed work, rather than trying to disguise these by adding superficial game mechanisms. Such extrinsic additions will likely not be sustainable over a longer period of time (Lepper, Greene, and Nisbett 1973). The enormous interest in gamification within the business world has, however, the ironic benefit of creating awareness about how improved working conditions can be financially justified. Gamification, done right, builds on the values identified in Self-Determination Theory (Ryan, Rigby, and Przybylski 2006), which are well aligned with what I propose here.

The actual implementation of the philosophy presented here does not differ much from most of what is considered as good design already today. Usability principles still apply, but are complemented with a direction that is not subjectively determined by either DDs or users: the levels of complexity and control in a tool are not determined by aesthetics, ease of use or learning, or any other instrumental value, but by what the user needs in order to be autonomous, which adds the dimension of appropriateness to the assessment of an artifact’s complexity and controllability.

This, as many other essential parts of this thesis, has been addressed in seminal HCI theory. The ideas that artifacts should be seen as tools and that they should give the user control have been around since the 80’s (Hutchins, Hollan, and Norman 1985; Ehn 1988; Norman 1986). How come we can still regularly observe effects of the opposite?

One explanation may be that these qualities, which essentially have to do with usability, are not considered as critical in the same way as, e.g., security problems, and are therefore not prioritized in procurements. Surely, most people would choose a tool with good usability, rather than poor, if they had the choice, but such choices are rarely given, and usability may be trumped by factors that can be appreciated more immediately. This, however, presupposes that there is a conflict between these specific aspects of usability and other factors, which leads to another possible explanation: it is not desirable for organizations to have autonomous members. One implicit purpose of ICT systems may be to function as tools for management. By demanding members of an organization to use specific ICT systems, it is possible to exert control over them.

At the Faculty for Science and Technology at Uppsala University, it has been suggested that all departments must use the same web content management system to create their respective home pages. Not everybody is happy
with the suggestion. The espoused purpose is to give a unified and professional impression to a visitor. However laudable such an ambition is, the effects of the suggestion are not limited to that. By deciding which tool is allowed, it is also possible to decide which type of content is allowed and who is allowed to edit it. Furthermore, it moves the responsibility for continuous development to a central authority, which does not experience the needs that specific users have.

A third explanation is that the offered insights are difficult to implement. What is the difference between regarding a computerized artifact as a tool and as a system? How do we determine the appropriate level of control? There is a risk that a resolution to the first question ends up being only a shift of terminology, but this is not my intention. Neither was it Norman’s nor Ehn’s. A tool view stresses that the user understands and is in control over what she does; that she can decide and act deliberately. This leads to an answer also to the second quandary: the appropriate level of control is determined by the user’s purpose and what competency in the specific task execution it is reasonable to expect from her.

8.5 Allowing Autonomy

If we want people to decide in a more autonomous way, we approach a curious dilemma: any attempt to stimulate autonomous thinking means exerting outside influence, which is inconsistent with the notion of the construct. Furthermore, most everyday decisions are usually routine matters—such as which bus to take to work, what and where to have lunch, how to handle time reporting—for which effortful autonomous thinking would be wasteful. However, it is also impossible to predict when we should employ it. Consequently, we should try the alternative approach, which is to remove conditions that reduce our possibility to use autonomous thinking when that is called for. It then becomes crucial to identify such aspects of artifacts.

8.5.1 Aspects That Reduce Autonomy

I will here account for seven aspects of artifacts that can risk reducing the possibility to think autonomously. These should not be seen as distinct categories: the subdivision is intended merely as an introduction to addressing users’ autonomy in design.

Confusion

If we cannot fully comprehend the situation that we are deciding in, it means that we need to rely on authorities or simply guess. This can happen if the artifact does not maintain and communicate a sufficient and consistent model of how it functions. As a consequence, the user cannot form a proper mental
model of the system, which means that she will not be able to appropriate it for her needs without the risk of causing errors. Neither will she be able to diagnose independently whether an error has been caused. The result is that she cannot take full responsibility for her decisions, as she is not able to use the artifact deliberately, i.e., be in control.

**Restriction**

If the user does not have the possibility to decide whatever she rationally concludes to be best, she is not autonomous. This may, or may not, be a problem.

If an artifact enforces one proper way of completing tasks, it has both positive and negative consequences, which need to be balanced. On the positive side, it makes it easier to do right and more difficult to do wrong. However, it is also restricting the user’s possibility to counter events that were unforeseen by the DDs, and synthesize different tasks in her work process.

Many administrative systems are not designed to handle exceptions from the normal. The result is that the user is either deterred from accepting and handling exceptional cases, or uses her cognitive resources for fitting the exception into the prescribed way of handling a case.

However, if there are less restrictions, there is a greater risk for errors. For this to be acceptable, consequences of misjudgments cannot be orders of magnitude greater than can be expected of an erroneous decision.

**Overautomation**

If the artifact does too many things automatically, it leaves for the user to do only the set of tasks that could not be automated. This reduces the user’s attention and her incentives and possibilities to learn how to use the artifact effectively (Bainbridge 1983; Stensson and Jansson 2014).

Examples of this fallacy are the autopilots that silently handle stress situations without the pilot knowing, thus creating a false feeling of safety, and nuclear power plants that have become so automated that only the ones who have worked there since the beginning, who have followed the evolution of the technology, understand how they work.¹

**Acting as Authority**

If the artifact allows the user to participate only in a subset of tasks, it gives the impression of a complex system that is beyond her control. This may reduce how much responsibility the user understands that she has, and may reduce the motivation to take responsibility. The effect of this, is that the user may adopt

¹This is based on private communication with a younger engineer at a Swedish nuclear power plant. When something breaks, he is dependent on that the system gives a warning. He then uses the blueprints over the facilities to locate what component to replace. It is not at all clear to him how the power plant really works and he worries for the day when the experienced engineers eventually retire.
an attitude of skilled incompetence; of blindly following practices that are not oriented toward any constructive purpose (Argyris and Schön 1996).

A human-machine system in which the user does fragmented work can also reduce the possibility to observe the consequences of her decisions. Without this feedback, an erroneous mental model cannot be corrected.

**Thinking for The User**

If the artifact does the qualitative thinking for the user, it makes her dependent on the artifact. In our current society, this is inevitable, and the ideal should not be that we become renaissance men. However, we need to be careful with which cognitive skills we need to maintain, and which are unnecessary. An interesting example of this is GPS navigation in cars, which for some people have become a necessity for being able to drive. As long as the device is there, it is just fine, but when it is not, the user is left without having developed the expertise to handle the situation.

**Oversimplification**

If the reason for using an artifact is to increase the possibility to take responsibility, its actual complexity cannot be hidden from the user. Instead, she has to be given means to handle it. Such means include both relevant information and stimulation of System II thinking.

Here a balance needs to be struck, based on what is reasonable in relation to the purpose of the user. If the levels of complexity and controllability are too low, the user is disempowered, and if they are too high, she will have difficulty to develop sufficient expertise. The effect is the same: she will not be able to take responsibility for her decisions. This is particularly important when it comes to expert systems.

**Emotions**

It may not strike the reader immediately that there is a potential problem in evoking emotions by artifact design. After all, much of User Experience is focused on creating positive experiences (Hassenzahl and Tractinsky 2006). Psychological research, however, suggests that emotions trigger the instinctive System I processes and suppresses the rational System II processes (Slovic et al. 2007).

A current trend within HCI is *Affective Computing*, within which one topic is Human-Robot Interaction. A typical question that is addressed within this topic, is how we can create more acceptance for robot assistants in, e.g., elderly care. One of the avenues that is being explored, is to make the robots sensitive to cues of emotions in the humans that they are interacting with so that they can respond accordingly. What we are dealing with here is pretense of empathy, intended to evoke positive feelings in the human. The ambition is to create assistive technology, which is laudable, but the evoking of emotional
response also means that the critical System II thinking of the human can be reduced.

This is not always a problem. Many times the intended effect of design is precisely to evoke emotions. Emotions are good for creating motivation, and are not harmful as such. After all, novels, movies, and games are very often designed to evoke emotional responses, and we are likely not harmed by that. The problem occurs when emotions are used to manipulate the reader, viewer, player, or user into making important decisions that would be different if the emotional stimulus was not provided. If we aim to maintain the user’s professional judgment, we need to be aware of how we address emotions in design. The challenge, thus, lies in creating emotional engagement without reducing the necessary attentiveness of the user.

All of these features can make it more difficult for a user of technology, either cognitively or by available actions, to diagnose problems in the artifact that they are using. Such artifacts do not enable poïësis, as Heidegger (1977) would put it, i.e., they do not allow the user to grasp the qualities of their essence, and they therefore can only become superficially ready-to-hand; they can extend the capabilities of the user only on the artifact’s terms. This does not necessarily mean that the user does not feel empowered, but if we look a bit closer at what this empowerment means, it is functionally fixed to the model that the artifact is designed according to. The user will not be able to appropriate it for her needs. In Designing for Appropriation, Dix (2007) suggests seven design principles to inspire designers to think of these issues. To summarize the points I have stated here, I have extracted a few important points from this paper.

allow interpretation—Don’t make everything in the system or product have a fixed meaning, but include elements where users can add their own meanings [...] provide visibility—Make the functioning of the system obvious to the users so that they can know the likely effects of actions and so make the system do what they would like. [...] expose the intention behind the system. [...] Instead of designing a system to do the task you can instead design a system so that the task can be done. (Dix 2007, pp. 28–29)

The primary consideration, however, should be to enable users to form proper mental models, both of the problems that they are facing, and of the tools that they are using. If they need to develop expertise in using the artifact, it is crucial to design for controllability and continuous learning. The artifact, thus, should never prevent users from utilizing and developing their skills, nor limit their development of understanding.

If mental models are important to consider in the design of software, how should they be addressed and can we see this applied in practice? More importantly, can we find support for Norman’s early claim that systems become enjoyable to use when they communicate their system models (Norman 1986)?
Let us, as an arbitrary example, take a smartphone email client named Gmail and look for examples of how the experience of using it relates to the formation of a mental model of it.

The product is most likely one of the most used smartphone applications out there. If we take a look at it with mental models in mind, we will find that there are some nice features that make it harder for the user to form a consistent mental model of the interaction with the application.

One example of this is that you can swipe horizontally with your finger to archive an email when you are looking at the inbox view (Figure 8.1a). This is convenient and makes it easy to maintain control over the inbox. If you, on the other hand, are reading an email, swiping horizontally moves you to the previous or next email. In this mode, archiving is done by tapping an icon in the top row, an action which moves you back to the inbox. This is procedurally inconsistent but the really critical scenario is if you have used the application’s search function to locate an email. In this list, which visually looks just like the inbox, the email is not archived when you swipe left or right, it is deleted! The only feedback that this has happened is a tiny text, which you have quickly learned to disregard after a few times of using the gesture in the inbox (Figure 8.1b). Still the application is pretty delightful to use. The swiping feels nice and the sober animation gives the impression of controllability.

The fact that you may accidentally delete emails is hidden from you until you unsuccessfully try to locate them again. When you cannot find them, you will not be able to diagnose why they are missing. A flawed mental model may thus not affect how enjoyable the artifact is to use, but it will affect your possibility to use it according to your intentions.
Now the interesting question to ask is where to draw the line—how much autonomy should the user have and how much can be handed over to artifacts? This question can be divided into seven, more detailed questions:

- Is it important that users do not make mistakes?
- Is it important that users follow regulations?
- Is it important that users can learn to use the system easily?
- Is it important that users have full control over what they do?
- Is it important that users can appropriate the tool to better fit their needs?
- Is it important that users develop expertise?
- Is it important that users can take responsibility for their decisions?

Every question should be followed by repeated use of why and why not. From the answers it is possible to gain insights into what the purpose of the artifact is. I will refrain from stipulating answers to them, as I believe that they should be discussed when designing a particular artifact. What is important is to formulate an explicit understanding of why autonomy should be allowed, to what extent, and how this is realized in an artifact. This implies formulating a design philosophy for each individual artifact, taking into account contextual needs.

### 8.5.2 Allowing Designers and Developers to be Autonomous

Autonomous thinking is necessary for solving problems that require thinking outside known sets of rules, and autonomy is necessary for being able to take responsibility. Given that I claim this to be important for users, it would be inconsistent to constrain the autonomy of DDs by forcing them to follow a rigid method that would ensure user autonomy. Rather, the possibility to be responsible should be granted also to them.

### 8.6 On Theorizing

Until recently I regarded theories as something grand and authoritative; ideas that were developed by extraordinary researchers as a life-time achievement. I am here thinking of theories like Darwin’s theory of evolution; Kahneman and Tversky’s prospect theory; and Festinger’s theory of cognitive dissonance. Most of such theories have emerged from hypotheses that have not been possible to falsify. These—at least the ones that have emerged in the natural sciences—go by the name scientific theories.

The fact that such theories have emerged from corroborated hypotheses naturally implies that most of them are rather sharply defined, since a theory that makes bold assertions about complex phenomena, for obvious reasons, is difficult to test unambiguously.
One could, however, argue that despite how rigid the logic in the theory is, there will always remain a certain measure of speculation, as novel insights depend on abductive reasoning. Popper wrote that “every discovery contains ‘an irrational element’, or ‘a creative intuition’” (Popper 2005, p. 8). Weick reinforces this view when he is advocating more theorizing:

“Regardless of how the trigger to theorizing is described, it consists of some description that can vary in fineness of detail, accuracy, and explicitness of assumptions which it incorporates.” (Weick 1989, p. 520)

What I call a theory here is not grand in any way. It is not even scientific, in the sense that has come to bear. I would like to advocate a slightly different perspective: to make theory useful as a description of what we, as humans, develop when we make sense of the world, rather than only regarding it as an argument to convince others that what is done is in line with what others have done before. I see great value in allowing ourselves to theorize more and be more generous about the content, although perhaps not the form, of theories (Sutton and Staw 1995; Weick 1995). A positive effect from extensive theorizing is that particular, situated, contextual, tentative theories are a way toward generating stronger theories, and by that, a way of advancing also research within the field (Weick 1995).

This liberal view of the concept theory is shared also by the organizational theorists Chris Argyris and Donald Schön. To describe the difference between what organizations claim that they do, and what they really do, they established the terms espoused theory and theory-in-use (Argyris and Schön 1996). In order for an organization to learn from mistakes that are committed by its members, the theory-in-use must be made explicit and be connected to the values and strategies of the organization. The intention with the here proposed protocol is to produce such explications about the use of artifacts, before these are created.

To argue for this view, I will give some definitions that hopefully convey more readily what I mean.

First, I define a theory as a model and an explanation of the the mechanisms of a phenomenon. It answers the questions what, how, and why. It constitutes the understanding of a situation and involves a representation, typically a simplification, of reality, with events and components of interest singled out.

This simplification is the model, and it is made under the assumption that what remains is sufficient to exhibit the actual dynamics of the situation. The old proverb, attributed to the statistician George Box: “All models are wrong, but some are useful”, is meant to remind us about this treatment. Ideally, models are concise descriptions of how a system works; how the significant components in the system interact and depend on each other.²

²Models can also be prescriptive. They can for instance be created as a blueprint for a system, like a UML diagram, or function normatively in social systems, in the sense that they prescribe
In the following I assume that a descriptive model can be identified, but this
is not always straightforward, especially when it comes to human systems.
Espoused organizational models will affect the way we describe our behavior,
but do not necessarily reflect how we actually behave. For instance, if someone
asked you who your boss is, you would probably answer according to the
espoused model, but would the answer be the same if you were asked whose
professional judgment you respect the most? If these answers differ, it will
become more difficult to describe who influences your decisions most, as the
boss can order you to do something, while, at the same time, the other person
will have influence on your judgment.

The purpose of identifying a descriptive model is here to allow theorizing
about what purposes and needs a solution to a problem should fulfill.

8.7 Theory of Use Protocol

My colleague Åsa Cajander has studied how systems developers in a large
governmental organization regard usability. The below quote is a reaction
from one of the interviewees, a systems developer, when suggested to work
more closely together with the intended users:

We are as good as they are at guessing! And I don’t buy all these things about
working with the users because they know how things work. Because I have
been talking to users and I know the rules better than they do since they are
used to doing things in a certain way. I, who don’t work with these things
say that I have read the rules and these are the rules. Then I can do it easier
somehow. I wouldn’t say that I am better than they are at saying what they
want. But still somehow we guess what they need and they are not better than
we are at expressing that need. Of course there are some things that they can
tell us, but I don’t think it is a good idea. (Cajander 2010, p. 71)

One could interpret this statement as reflecting an over-confidence bias of the
developer, or see it as indicative of an einstellung effect in the users. Both of
these analyses likely have some truth to them. However, what I wish to point
out in this quote is that the developer is expressing exactly the mindset that
the below suggested protocol is intended to structure into constructive, rather
than confrontational, dialogue. The developer may very well be right that the
users are not better at expressing their needs, as that is not an integral part of
their training. Nevertheless, they would likely realize what they do not need if
he presented his educated guess to them.

roles to individuals. An example of the latter is a formal organization. However, if one would
study the dynamics of it, the espoused organizational model is not necessarily the most accurate
description of the actual dynamics. It serves more the purpose of communicating organiza-
tional values, while the model-in-use: the actual values, practices and structures, are often more
pragmatic (Argyris and Schön 1996).
Figure 8.2: When designing a new system, DDs take the current system and practice, as well as the user’s mental model, in account. The mental model that the user then forms of the new system is affected by her mental model of the system that is replaced.

Also, it becomes clear from the quote that the DDs basis for designing the system is the rules that are to be followed; not the autonomy of the users, and not the purpose for which the rules were created.

Figure 8.2 shows a schematic model of how a new system is designed in current practice. It is intended to highlight how the currently used system affects also the design and apperception of a future system, if this is not explicitly addressed. DDs take the current system and practice, as well as the user’s mental model, in account in the design of a new system. The mental model that the user then forms of this system is affected by the mental model that she had of the system that is replaced. This is likely inevitable and should, thus, be taken in account.

The reason for designing a new ICT system should, however, be nothing else than that the current system—regardless if it is implemented digitally or not—does not satisfy the requirements of the user or the organization in which it is used. The process of developing a system should neither depend on whether there already exists an ICT system that is to be replaced. Content-wise, the difference may be great, as practices may have to change dramatically when ICT is introduced into a previously “analogue” organization, but in both cases there is a desire to improve some aspects of an organization’s value system. The analysis should, thus, focus on the needs of the organization, including the needs of the people that constitute it, rather than on what can be learned from the old system.

An optimist would think that reform initiatives are driven by a clear idea about what is working badly and how to improve it, and most important of all: what would work really well. However, judging from the systems that I have encountered both professionally and privately, this does not seem to always be the case. Bad systems have been replaced with even worse ones.
What I advocate here is not a revolutionary new way; the process is several hundred years old and it is used implicitly, every day, by various domain experts. What I attempt is to formulate a protocol out of something common-sense, something that is already state of the art among those DDs that produce good designs. Although common-sense, it is still counter to what we usually do, and thus not something that we engage in without effort. It is the falsification, rather than verification, of hypotheses.

By this, I want to approach a formalization of the magic that happens between problem discoveries (e.g., ethnographies or strokes of genius) and design.

One of the benefits in focusing on falsification is that a possibly flawed or incomplete conception of the situation can be made apparent before the actual design has been made. Another benefit is that it counters the human tendency to look for evidence that confirms what we think is right; our confirmation bias.

In order to enable falsification, I deem it necessary to first shift the roles of DDs and users slightly. Instead of the DDs treating users as objects of study, users should be empowered to assess and falsify the DDs’ understanding of them. Scrutiny needs to be reversed. As the users are no longer the objects of study, but the subjects that are studying, the resulting process becomes truly user-centered. This may at first seem like an ungrateful degradation of DDs, as their decisions will no longer be dominant. However, it means that they will have less responsibility for making correct initial assumptions about users, but more responsibility to act professionally. I thus argue that the effect is the opposite: by reducing the responsibility for guessing what the users need, DDs gain the opportunity to make full use of their creativity and their specific expertise.

Second, design proposals, prototypes, and even the resulting artifacts need to be considered as hypotheses, rather than as solutions. Apart from making explicit that these are not definitive—that they are malleable—this also creates an expectation that they are constructed according to a theory. Presently, the equivalent to such theories, i.e. requirement specifications, seem heavily based on functional requirements with mostly unarticulated assumptions about users and the use situation. The shift of perspective puts focus on acknowledging and making explicit the knowledge, presuppositions, and ideas that DDs have about users, their situations and their practices.

I call such an explication a Theory of Use and it comprises the DDs’ understanding of the users and the use situation. The protocol resembles Grounded Theory Methodology (Glaser and Strauss 1967) in that it tries to mimic the way humans naturally make sense of their environment, and serves the purpose of generating a theory. The important difference is the focus on falsification of assumptions, rather than on categorizing data.

The value of the proposed protocol can be summarized in four main claims that I will detail below:
1. It promotes thoughtful design, founded in agreed-upon arguments.
2. Scrutiny is reversed and, by that, the process becomes truly user-centered. Instead of DDs treating users as objects of study, users are allowed to assess the assumptions that are made about them.
3. It limits the time that is invested into constructing the wrong solution, as users can react already when the DDs account for their understanding of what they are designing for.
4. It is formative and lightweight. When a Theory of Use is accepted by the intended users, DDs are liberated to focus their expertise on developing hypotheses, i.e., design proposals, that are then, again, evaluated by the user.

The intention of the protocol is mainly to facilitate constructive dialogue between DDs and users, but I also believe that an adoption of this way of thinking about design will contribute to making HCI a more practically relevant area of research, and paradoxically also a more scientific enterprise.

One premise for it to work is that the DDs who are involved are skilled and confident in what they are doing, in the meaning that given enough information about what they are supposed to design, and what they are supposed to design it for, they will do it well. Naturally, this is an obvious point of failure, but with sufficient dialogue, the risk will become apparent early in the design process.

If DDs are to be explicitly held accountable for their designs, they need to be given all means possible to completely understand the problem that they are solving; they need to be allowed to form a proper mental model of the use situation and the users’ needs. They also need to enjoy full freedom to make design decisions founded in rationality rather than in imposed rules. Most importantly, the protocol needs to be forgiving to errors, both because stress inhibits the ability to employ System II reasoning (Kahneman 2012), and because errors that are revealed are falsifications of current understanding: by exposing misunderstandings, thinking can be corrected.

Everybody has preconceptions and it is better to make these explicit, rather than letting them influence tacitly. Also, when they are made explicit, others can react to them. This is an activity that everyone involved in the design process should participate in. The effect is that differing perspectives and opinions are visualized and the team will be able to have a dialogue about what would otherwise be left unspoken.

The protocol means that evaluation is refocused, from a mean to assess design, to a way of improving the mental models of DDs. The difference may seem subtle, but it is necessary for being able to acquire expertise: if DDs cannot trust their ability to design good artifacts—if evaluation always trumps expertise—they will not be able to form the mental model that is necessary for understanding the problem, and they will not be able to make well-founded predictions about future use.
The steps, and the respective responsibilities of DDs and users, in the protocol are presented in Figure 8.3. The outcome is a hypothesis of an artifact and a document, a Theory of Use, that explains it. Below, I detail important aspects of the protocol.

8.7.1 Design is not Experimental Psychology

As I am to advocate hypothesis testing, I find it important to be clear about what I do not mean by that. The classic ideal for hypothesis testing within the field of HCI is a psychological experiment. The definition of usability gives three dependent variables: efficiency, effectiveness, and satisfaction. The independent variable that is to be manipulated is the artifact, most commonly some aspect of a piece of software.

As HCI is a practically oriented discipline, there is a strong assumption in such experiments that the dependent variables in the experimental setup are representative for successful task completion in a natural setting. The corre-
lation has not been unanimously proven and, consequently, many researchers have abandoned laboratory experimentation to instead employ observations in more natural settings (Shneiderman 2008). There is, however, also in such an approach a risk, namely to assume too much generalizability from the test cases. One may object to such a remark, as the artifact that is used in the experiments is the one that will be used in everyday practice, hence having identical cues as in the natural setting. This is true in a sense, but how often is it really the case that the observed situation is like the ordinary? The most important part of working with computers is often cognitive, i.e., happening in the brain of the user. Situated observations typically use predefined, allegedly representative, tasks that a user is to perform. These will reveal some aspects of the artifact’s usability, but will they generalize to the full set of tasks for the user? Would the experience of completing the very same work task be the same without the detailed instructions that are common in usability studies?

Although controlled studies are good for identifying obvious usability problems, the actual use situation, for which the artifact should be designed, is captured better by ethnographic methods; by patiently observing, without giving any instructions. The work of a designer is predominantly epistemological. It is about discovering what goes on in a situation and propose an artifact to support a user in her struggle to fulfill her needs.

8.7.2 A Hypothesis

There are several possible ways to define a hypothesis. What all of them stress is, however, that a hypothesis is tentative. Within psychology, it is common to regard hypotheses as falsifiable statements: “If I manipulate the independent variable X, then the dependent variable Y is affected like this”.

This is also the most straightforward way of doing hypothesis testing of software and artifacts: we define a task, use case, or scenario, and set up a number of hypotheses along the lines of “users of software A will solve task X more efficiently than users of software B.”

Here the artifact is not the hypothesis in itself, and is instead used as an independent variable in one. This type of experiment, if done properly, yields information about how appropriate some artifact is for a particular task, which is very useful information. However, it is often impractical to set up this type of studies, and artifacts are rarely used in controlled environments, by people who devote their full attention to the task at hand.

This type of experiment tests the task-tool fit, but does not address the purpose-tool fit. Actual practice is usually difficult to control, and there are constant interruptions and continuous changes of plans. The idealized situation, which is used in most controlled experiments, is rarely even what is desirable, if one believes that motivation, inspiration, and serendipity are important factors. Over the course of an experiment, it is not unreasonable to suspect
that the test subjects maintain motivation to a greater extent than would be
the case in every-day practice, they rarely need to be inspired to complete the
given tasks, and serendipity would definitely ruin the experiment.

I would like to address a tacit assumption that is made in the above type
of statements: why do we believe that the manipulation yields that particular
effect? The answer is, of course, that there is a theory behind it, that predicts
a certain outcome. To stress this, I suggest two alternative, almost identical,
formulations of what a hypothesis is.

1. A hypothesis is a reformulation of a theory so that it can be tested.
2. A hypothesis is a specific prediction that is motivated by current knowledge,
   i.e., theory.

These particular wordings are chosen because they clearly connect a hypothe-
sis to a theory, and suggest that the reason for making hypotheses is to be able
to develop a stronger theory.

An Artifact as Hypothesis
If we now consider the human mind, we can put forth the idea that: connected
to every deliberate intervention, there is always an implicit theory—a mental
model. It may be flawed, with the consequence that I make the wrong interven-
tion, but there is always an explanation, tacit or explicit, to why I intervened in
one way or another. The intervention is, consequently, a hypothesis about what
is correct to do—if I do this then that happens—based in the mental model that
I have about the situation. We could also imagine that the definition of an ar-
tifact, as intentionally made by humans (see Section 4.1) assumes deliberate
interventions. The consequence is that artifacts can be seen as materializations
of hypotheses. If we evaluate the formulations above by replacing hypothesis
with artifact we get the following suggestions:

1. An artifact is a reformulation of a theory so that it can be tested.
2. An artifact is a specific prediction that is motivated by current knowledge,
   i.e., theory.

Although both of them make some sense, the first one is rather abstract, while
the second one is what we usually want to accomplish in design. What I
essentially propose is to understand and scrutinize artifacts in the same way
as we understand and scrutinize hypotheses, as specific predictions. By the
associations that the latter is loaded with, the hypothetical properties of the
former are revealed. Artifacts are hypotheses of what fills a purpose for a
specified user (Woods 1998).

Much like a theory, an artifact embodies a description of a use situation,
though with the constraint that the artifact can only represent it, but not convey
the rationale for its functioning. The utility of an artifact is proven in use, but
if it is not the optimal solution for some given situation, the existing design will not necessarily help DDs to improve it to better serve the users’ purposes.

If artefacts embody theory, however, they do not encode it, and if they occupy a point in a design space, they do not highlight the salient or fruitful dimensions of variation that space offers. (Gaver 2012, p. 944)

This makes an artifact resemble a hypothesis more than a theory—the artifact is a prediction about purposefulness, but it does not explain why it is purposeful. In order to improve the artifact, DDs needs to have an idea about this, i.e., a theory. From a theory, they can form other hypotheses.

The Null Hypothesis, H0

In customary hypothesis testing, it is often talked about the null hypothesis, i.e., that manipulation of the independent variable did not lead to the predicted effect. The purpose is to establish criteria for falsification.

When testing a pill, it is the case that no medical effect could be determined. Placebo is used to control to what extent belief affects the result. What is the null hypothesis when testing tools? For the hypothesis described above, it would be that users of software A did not solve task X more efficiently than users of software B. Would it be possible to formulate this when the artifact itself is seen as the hypothesis?

Here we need to realize that the purpose of the null hypothesis is to make it possible to reject a hypothesis; to make explicit when it is not supported by evidence. If the artifact is considered as a hypothesis in itself, which is implicit also in the pill testing case, the null hypothesis would simply mean that the theory was not corroborated. The Theory of Use should, thus, include an account of when it is not supported by evidence. This fills the role of the null hypothesis.

In the case of an incremental improvement to an existing artifact, it may be easiest to use the version of the artifact that is to be replaced as null hypothesis, which makes comparisons rather straightforward.

In the case of introducing a tool for work, it is slightly trickier. If the artifact is introduced and workers are forced to use it, we will definitely have some effect. Here we need to make use of the qualities in a good solution, which are described in the Theory of Use. These qualities need to be formulated so that it is apparent when they are not fulfilled.

In cases where an innovation is introduced to the market, we can imagine two approaches. The first is simply that status quo prevails; that users do not adopt the artifact. This binary approach does, however, not yield much input for improving it. More often it is tested on potential users before market introduction. In such a case, one has to include in the qualities what purpose the artifact is to fulfill—what problem it is a solution to, as Neil Postman would put it—and form a hypothesis for how this is solved without the artifact.
Although this may seem abstract and difficult to define exhaustively, it is a valuable step to take, since if the purpose for creating the artifact cannot be identified, there is a risk that there is none. By this, I do not claim that such an artifact is necessarily useless, as it is still useful for exploring possible futures.

**Falsification of an Artifact**

It may seem constructed to formulate a specific null hypothesis in order to enable falsification of an artifact, especially if there are multiple reasons for falsifying it. What is important is to be alert to when the artifact is not a correct prediction. It is falsified if it does not fulfill its intended purpose. Therefore, even if a specific null hypothesis is not formulated, it is important to establish, in advance, what the purpose of the artifact is.

Falsification of an artifact means exposing it to its intended users to get their feedback. However, it is important to neither accept positive, nor negative reactions at face value. Falsification should not be based on instinctive reactions, but by carefully reviewing whether the artifact fulfills the qualities that are stated in the Theory of Use. Through this, the falsification serves a formative purpose of adjusting the Theory of Use, i.e., the DDs’ mental models, so that the next hypothesis, if necessary, will be an improvement. It is a natural part of an agile, iterative process to revise initial assumptions. Consequently, what is important is to understand why the hypothesis is rejected so that this aspect can be addressed further.

**8.7.3 Find Meaningful Idea**

For design to be relevant, there first has to be a problem that requires solution, or a need to fulfill. The problem or need can, however, be something that we are not aware of yet, which is the case with much of consumer technology. Twenty years ago, the number of people who needed to browse the Internet on their phones was probably smaller than today.

The type of problems that do not follow logically from what we already know cannot be identified scientifically. Those require the ability to notice flaws in the current situation and imagine different futures. In this process, knowledge about the human mind—informed by psychology and philosophy—can guide the focus, but the most important ability is, however, the elusive one called creativity. I will refrain from trying to define it, but what is clear is that it presupposes abductive and autonomous thinking (Sheldon 1995). By rational, logical thinking processes, we can discover new consequences of current knowledge, but we will not be able to reach beyond it.

However, the principle of autonomy can be used to deductively identify a set of problems. If we assert that people want to be autonomous, we can observe the status quo and pay attention to how autonomy is traded off for other values. We could, for instance, hypothesize that people buy an expensive car because that gives them freedom and control.
8.7.4 Theorizing

When a meaningful idea has been identified, it needs to be specified. Following Popper's reasoning, ideas are made concrete by means of a logical deduction process.

From a new idea, put up tentatively, and not yet justified in any way—an anticipation, a hypothesis, a theoretical system, or what you will—conclusions are drawn by means of logical deduction. (Popper 2005, p. 9)

When the intention is to design an artifact so that it satisfies users' needs, logical deduction means to work systematically with explicating and exposing the DDs’ mental models of the users and the use situation. This explication is the Theory of Use. The tentative idea that DDs have, is concretized and formulated in statements that are possible for the intended users to falsify. One of the underestimated activities that happened in the UTOPIA project were the so-called quality workshops. Pelle Ehn mentions it only in passing:

Another activity in this phase consisted of so-called 'quality workshops'. In these workshops graphic workers, graphic designers, and journalists participated in discussions of good use quality of graphic products. (Ehn 1988, p. 332)

In the Theory of Use protocol, such quality workshops can contribute to developing an understanding of what the purpose of creating an artifact is. By discussing qualities in a good solution with users, the DDs come to form an initial understanding of what the problem is that they are to solve. This is similar to the workshops that are detailed by Hardenborg (2007) and Hardenborg, Kavathatzopoulos, and Sandblad (2007). Focus should not primarily be on how to accomplish tasks, as the users’ understanding of this must be assumed to be shaped by how things are done currently, but on why. If the DDs understand why they are designing an artifact, they are liberated to use their expertise to suggest how it is best done.

The Theory of Use should also state what assumptions the DDs are making about users, technology, and relevant factors in the environment.

In this part of the protocol, the DDs initial responsibility is for forming an understanding of what the problem is that they are to solve. The users’ responsibility is to be explicit about the practices and the purposes that they can identify in using the artifact.

Explication of Implicit Theory

[...] the likelihood of a solution is determined in part by the way the environment is represented or perceived. [...] Theorists both choose the form of the problem statement and declare when their thought trials have solved the problem they pose, a sequence that resembles artificial selection. Theorists are both
the source of variation and the source of selection. People searching for niches also may choose the form of their problem statements, but the environment declares which of their trials, if any, are solutions. This latter sequence resembles natural selection because the source of variation differs from the source of selection. Despite this difference, solutions in both cases are more likely to be discovered where the representations are fuller. Whether the problem is to find an explanation or a competitive advantage, fuller descriptions suggest a greater number of possibilities. (Weick 1989, p. 520)

When the DDs have formed an understanding of the purpose, they formulate a Theory of Use. I propose a minimalist structure for the report: first the identified problems are introduced, followed by an account of what the DDs consider to be qualities of a good solution and the assumptions that are made about users, technology, and the use situation. The latter represents a model of the human-artifact system.

References to existing theories should be used to support reasoning, not to verify claims. Incentives should be clarified and put to the fore, including expectations on users and, if relevant, the business model around the system that is being created. This is important because it frames the possibilities that the DDes have, even more so if it is left unstated.

For instance, if a chain store manager is dissatisfied with how cashiers perform, and wants to purchase a system that forces these to uphold a certain pace, it is likely based in a conception of the cashiers as lazy, insubordinate, or indifferent. If this actual assumption is left unstated, the DDs lack a piece of the puzzle, which could lead to the development of the wrong type of system. However, if it has been explicated, there is a concrete assumption to address, falsifiable by the users, which gives the opportunity to identify a solution that is better for both cashiers and management. This perhaps sounds idealistic, but also crude market logic would reach to the conclusion that workers who are productive and satisfied are better for the business than workers who are just productive.

There is remarkable power in matching the qualities of the available material with the qualities of a good solution. Often a first hypothesis is immediately given. The difficult task is to counter the functional fixedness of the material, to realize the affordances that it really has and not only those that are familiar. Another is to realize the qualities of use. Here, the principle of autonomy may help. Assuming that users want to be as autonomous as possible, how should the system be designed?

8.7.5 Let the Intended Users React

An objection against DDs formulating their mental models as a Theory of Use could sound something like: *Even if you tell the users: You will use the imag-
ined system like this, they would not know how they will use it. They would also be guessing, especially before they know how the system will actually be.

This is absolutely correct, but the Theory of Use is more about what to do and why, than about how to do it. The how is up to the DDs to figure out. It is the DDs’ hypothesis of what the users need, and after a design proposal is prepared, it is again time to let the users react. The protocol simply puts focus on the most important issue, which is to improve the chance that the DDs understands why, for whom and for what context, they are designing. The assumption is that if the DDs knows everything they need to know, they will be able to design a system that better fits the users’ needs than what the users would if they would do it by themselves.

In order to avoid anchoring bias, i.e., undue attachment to a reference point, users do not initially review design hypotheses, but the theory that will lead to the design: The Theory of Use.

Theorists are usually pleased when their assumption are disconfirmed whereas nontheorists are worried when their assumptions are disconfirmed. A disconfirmed assumption is an opportunity for a theorist to learn something new, to discover something unexpected, to generate renewed interest in an old question, to mystify something that had previously seemed settled, to heighten intellectual stimulation, to get recognition, and to alleviate boredom. However, a disconfirmed assumption is a problem for a nontheorist because it suggests that past experience is potentially misleading as a guide for subsequent action and that coping may be more difficult. (Weick 1989, pp. 525–526)

Falsification means to subject a hypothesis to test in the situation that it concerns. This is the only mean we have to develop our thinking. In this phase of the protocol, the assumptions in the Theory of Use can be seen as hypotheses. The implicit prediction is that the intended users will accept them. By this, the DDs put their understanding of the problem up to test and can learn if they have understood the problem correctly. What the users will assess is how well the DDs have understood their situation, their abilities, their needs and purposes, and their values.

The users can, however, not be expected to have reflected explicitly on these issues, so it is very important that the statements in the Theory of Use are formulated contextually, in a language that the user understands, even if they were derived from high-level abstract theorizing. For instance: “The task requires holistic knowledge of the logistic processes in the system and its current state” can be reformulated as, for instance: “To do your job well, you need to know where the trucks are and what routes they can take”.

Peculiarly enough, when the users accept a falsifiable statement about their interests or needs, it immediately takes the form of an argument for a yet not conceived solution: “To do my job well, I need to know where the trucks are and what routes they can take”. An important effect of this is that it counters
the human tendency to look for arguments that justify a decision after it has been made; the confirmation bias.

The Theory of Use should be revised and again subjected to scrutiny by users until it is accepted in its entirety. This phase of the protocol is the most important one, as it develops the DDs’ understanding of what is the problem to be solved and what a good solution satisfies.

This should be familiar to researchers and practitioners who are accustomed to traditional user-centered design, since it naturally involves interaction with the intended users of the system that is being developed. At the same time, I would argue that the perspective in the current proposal is a bit different. I agree with the traditional approach, that the users’ needs and abilities should be the guiding light of systems development, but I claim that there are strong reasons for putting the DDs’ apperceptions of users’ needs and abilities under scrutiny. DDs enter the problem solving process with preconceptions. Anchoring and over-confidence biases affect what is observed and understood. Explication of mental models is a powerful tool to counter this. This means that users are not necessarily involved in the designing of systems but instead function as reviewers of the designers’ work.

8.7.6 Make Hypotheses

When the users are confident enough that the DDs understand their problem—when they cannot identify flaws in the Theory of Use that is presented to them—it is finally time to start creating alternative designs. These should be treated as hypotheses rather than as solutions. The importance of this attitude, and the value in making concrete design hypotheses, is illustrated in a quote from a participatory design project:

> Decisions were naturally backtracked and revised, often as a result of the impression immediate visualisations in the prototype made. This was an efficient learning process. For example, one driver suggested that safety critical information that was presented as text on the blue background should have the color red. When the text was changed to red, everyone immediately reacted and realized that this text would be impossible to read. A corresponding decision for the future work was that the color red only should be used for warnings, and red text was never to be displayed against a blue background. (Olsson and Jansson 2005, p. 161)

As the hypotheses are put up for falsification—by the users for which they are designed—the DDs are relieved from the responsibility to make decisions on behalf of users. When they are more free to use their creative problem solving skills to address the assumptions and qualities stated in the Theory of Use, novel, innovative ideas can be evaluated. The general formulation of hypothesis testing in this phase of the protocol is the claim: “This artifact solves your problem”.

115
Also, if a design is treated as a hypothesis, we realize that observations that lead to immediate adjustments of the design means the type of spurious hypothesizing, described in Section 3.1. The adjustments are made to fix observed problems, which is a natural tendency of conscientious practitioners, but this refinement process should not be performed without developing the explicit understanding of the problem, i.e., the Theory of Use. After all, this should represent the mental model that the DDs have about the use of an artifact. A design is a hypothesis of a solution, based in the theory.

What should be done instead, is to focus evaluation on how well the hypothesis follows from the Theory of Use. When hypotheses are presented to users, it will likely become clear that some assumptions were not accounted for and that others turned out to be wrong, which means that the Theory of Use needs to be revised. Hypotheses work as probes into possible futures, to inform the understanding of both DDs and users.

It is also likely that misinterpretations occur in the forming of hypotheses from the Theory of Use. If users point out that a hypothesis is wrong, that says something about the DDs mental models of the problem that they are to solve. This requires careful examination, to determine whether the Theory of Use needs to be revised, or whether the DDs have made mistakes when deducing hypotheses from it.

If the error is in the mapping from theory to hypothesis, the Theory of Use may have to be clarified. If, on the other hand, the error is not found in the mapping, the Theory of Use needs to be revised according to the newly identified assumptions and qualities. To identify such errors, it is valuable to be able to point to the agreed-upon assumptions and qualities. As the Theory of Use is revised with the newly acquired information from the hypothesis testing, hypotheses are easy to discard.

Considering that we are here testing hypotheses, we should find a way of either avoiding to affirm the consequent (see Section 3.1), or being able to repeat the test until we can have some confidence in the hypothesis. The latter approach probably has to be rejected, since for many types of artifacts that are produced within the field of HCI—especially for expert systems—it is often not possible to recruit enough credible test subjects to make it meaningful to perform statistical analyses of the type that is common within medicine and psychology.

A more intriguing approach is for the DDs to deliberately include assumptions that probe outside of their current understanding. This means that they are trying to falsify the assumptions that they are not making, which can be especially useful if the DDs start to suspect that they are narrowing down on the wrong track. By including qualities and design hypotheses that explore distinctly alternative solutions to the problem, they can increase their understanding further.
8.8 Risks

Even if autonomy is regarded as a necessary value, there are several risks following from increasing it, especially if it is done without solid theoretical foundation. Each of these will turn to actual problems if they are not addressed appropriately. Below, I account for three of these.

8.8.1 Stress

Being explicit about one’s own understanding—thus accountable for it—may result in increased stress. This is particularly true if the motivation is not to increase the chance of designing the right solution but only to expose inabilities of DDs, which could very well be a motive for some managers (Kruger and Dunning 1999). For the protocol to work, DDs and users need to realize that the groups have distinct responsibilities: for the DDs to clearly express their current understanding, and for the users to assess and rectify this understanding. The responsibility for correctness is thus shared between the groups, which is only effective if the DDs base their later hypotheses on the mutually agreed-upon assumptions and qualities that are stated in the Theory of Use.

8.8.2 Errors

If there are no strict rules to ensure that correct decisions are made, one may fear that a lot of errors may result. It may thus be tempting to reduce autonomy for the sake of correctness. For some types of work, this may be the approach to take, but if the work involves any amount of uncertainty, or demands creativity, it will likely be suboptimal. The proposed protocol is not devised to reduce errors, but to make it easy to recover from them. If a hypothesis proves to be wrong, it is being resolved either by revising the Theory of Use or by identifying weaknesses in how the theory has been operationalized as hypotheses. Both types of errors yield information about how DDs understand the problem that they are working on.

8.8.3 Biases

When DDs have formed a Theory of Use, they are expected to revise it when users react negatively to its statements. This reduces the risk for biases, especially the overconfidence, anchoring and confirmation biases. However, something that is not explicitly addressed in the protocol is the biases of users. The eINSTELLungs effect and Maslow’s hammer are highly relevant: users may often prefer to do things in the way that they are used to. They are also likely to anchor in initial qualities of a good solution for their problem, and be biased to judge subsequent information according to it. Awareness of biases, however,
reduces the effect of them (Arnott 2006) so it is important to communicate this risk.

8.9 Related Work

Here I will give brief accounts of concepts that have either contributed to the current proposal, or can be easily confused with it.

8.9.1 Cognitive Work Analysis

Cognitive Work Analysis (Rasmussen, Pejtersen, and Goodstein 1994; Vicente 1999) is a framework for systematic and holistic analysis of work and the environmental constraints that shape it. In Kim Vicente’s description of the framework, autonomy of the user is a central concern. The main argument is that a user needs to be prepared for unexpected events, and must, therefore, be in full control over her working environment. In my analysis of the same problem, I have come to add the moral dimension of how humans should be regarded. Autonomy is that ability, which distinguishes humans from machines and what earns us the right to be treated as ends in ourselves.

The ambition with Cognitive Work Analysis is to establish a formative work analysis framework that helps us specify the design attributes that computer-based information systems should have to satisfy safety, productivity, and health goals. (Vicente 1999, p. 62)

By formative, he means a work analysis that does not assume humans to function exactly as assumed, and that goes further than just describing how humans operate in specific settings. In my wordings, he aims at finding a theory of good work, founded in the conditions of a situation.

Vicente stresses the importance of identifying and explicating environmental constraints, in order for a user to avoid doing harmful things. This is a leap forward from earlier methods that were mostly focused on identifying the set of legal actions, thereby constraining the user unnecessarily.

The ambition with the present Theory of Use protocol is similar to Cognitive Work Analysis. The most interesting difference is to be found in the idea of switching the role of the observer—from DDs to users. This reduces the need for the structured formal analyses that Cognitive Work Analysis involves. It is, thus, possible to apply also in smaller project, in which available resources do not allow the rigidity of Cognitive Work Analysis.

8.9.2 Other Concepts

It is a fundamental insight to the field of HCI that the needs and interests of users are what should determine the design of artifacts. Several ways have
been tried to formalize this. Here I will go through some of these and point out what I believe to be the reasons why these have not had the expected success.

**Sketching and Prototyping**

Although most people probably associate sketching with drawing figures on paper, it is not necessarily only that. For some years now, the term has been used also for software that is intended to illustrate artifact functionality. What is important in a sketch is that it allows quick initial evaluation of a thought (Buxton 2007).

Prototyping implies to create a seemingly functional instance of the intended artifact. It is a common approach for proving concepts or evaluating how well interaction works in practice. However, if the intention is to glean information about users’ interests, prototypes are probably not suitable. Apart from the fact that it is relatively expensive to develop them, they may constrain the imagination of the user. A prototype usually has a limited set of designed affordances, to evaluate a specific way of interacting with it. Rather than focusing what she really needs, the user is set to evaluate how well the prototype functions. This is one of the strongest arguments for using sketches, rather than prototypes, when probing for information about how to design an artifact. A sketch signals clearly that it is tentative and disposable (Greenberg and Buxton 2008).

**Personas**

Personas are fictive individuals that are created to represent the intended users of a system. The purpose with personas is to assemble all the various pieces of information about real users into a few archetypes that can then be used as targets for design. They are strongly advocated by some usability practitioners:

> Just as physicists have created models of the atom based on observed data and intuitive synthesis of the patterns in their data, so must designers create models of users based on observed behaviors and intuitive synthesis of the patterns in the data. Only after we formalize such patterns can we hope to systematically construct patterns of interaction that smoothly match the behavior patterns, mental models, and goals of users. Personas provide this formalization (Cooper, Reimann, and Cronin 2007, p. 76).

The benefit is that they are easy to understand. They appear like people and they stimulate developers to think about the characteristics of those individuals when developing the system. The main problems with personas is that they are by design fictive, which means that claims made about them are not falsifiable.

As they also exploit prejudice, it can be a problem. If a developer asserts that a particular persona would like to have certain functionality in the system, then it could technically be equally possible that he would have said the exact opposite. Historically, personas have been successfully used in advertising
where the concept in a sense is self-sanitizing: if the construction of the persona is flawed, the advertisement will not succeed. The same argument could be used for artifact design, if the use of personas was restricted to consumer products. This is, however, not the case. When personas are used to build systems for which the ones who procure the system are not the same as the ones who have to use it, there is a great risk of personas functioning as straw men. Individual objections against the representativeness of a constructed persona are muted, as it is not even meant to represent individual users.

**Requirements Engineering**

Traditional Requirements Engineering is probably the most common tool to communicate what kind of system should be constructed. The way the requirements specifications are worded varies from very informal to very formalized.

A good requirements document is unambiguous, complete, correct, understandable, consistent, concise, and feasible. Depending on the customer-supplier relationship, the requirements specification can be part of the contract. (Paetsch, Eberlein, and Maurer 2003, p. 311)

One problem with requirement specifications is that they tend to become heavily weighted toward functionality. The documents thus end up rather abstract, often stating requirements without justification and without concrete addressing of non-functional aspects.

The difference between a requirement specification and a Theory of Use is in the focus. Where requirement specifications to a large extent focus on functional requirements and features in a system, the Theory of Use contains a model of the situation in which the artifact is to be introduced and focuses on qualities, in terms of fulfilled needs and goals, in a good solution.

**Concept-Driven Interaction Design and Annotated Portfolios**

Stolterman and Wiberg (2010) advocate advancing theory about artifacts in what they call *Concept-Driven Interaction Design* Research. In brief, the essence of the idea is to recognize paradigmatic designs as carriers of knowledge.

In interaction design research, this means that a new concept design has to express the underlying theoretical concepts it is supposed to manifest in a true and authentic way. (Stolterman and Wiberg 2010, p. 105)

For research about novel interaction devices, this could be a suitable approach. What it requires is that the device has one clearly isolated functionality that the design comes to embody. However, one problem in representing theory in a solution is that it reduces our cognitive ability to “unknow” (Stroop 1935)
the solution. Another potential problem is that theorizing is done based on
the design concept. The positive aspect of this is that artifacts can be used
as probes into possible futures, but it can also lead to case-based post-hoc
rationalization if applied frivolously. Theories should not be something that
are constructed after the result is produced, but rather before.

Annotated Portfolios (Gaver 2012) is a similar approach, in which important
design decisions are annotated with arguments to justify why those decisions
were made. This is similar to what I advocate, with the difference that I claim
that the arguments should precede the design so that they do not risk becoming
a rationalization of already made decision. In this aspect the proposal here is
more normative, demanding a shift of perspective for how design is produced.

8.10 A Note on Quantitative Data

I do not explicitly advocate quantitative methods, like statistics, but neither
will I argue strongly against using such. Whether these are useful or not,
depends on what type of claims one wishes to make. Even though I use the
scientific method as inspiration in the following, I care more about how much
information a method yields than about the form it takes. What is scientific is
not determined by whether numbers are involved or not, but by the resulting
understanding and generation of theories. In special situations, staged by an
experimenter, quantitative methods have worked well for exactly this purpose,
but they have not always generated knowledge that is applicable outside of the
well-defined context. Quantitative data is the answer to certain questions, but
it would be foolish to believe that it is what constitutes science, and that it can
answer everything that scientists are interested in.

The situations in which this approach is just not the way to go are numer-
ous. For many people, not even routine work means that a set of events occur
in the same sequence every day, but rather that a set of unsurprising events
occur stochastically, with unsurprising frequencies. This poses problems for
controlled experiments in HCI, as the number of independent variables in-
creases, and with that, the number of participants required for a statistically
valuable result multiply. It would also be counter-productive to throw away
that information, which cannot be enhanced meaningfully by aggregation, in
order to fit data into a scientific ideal.

In The Sciences of the Artificial (Simon 1996), Herbert Simon uses the pol-
icy making for automobile emission as an example of a problem in which
quantitative estimates were produced, but contained such ambiguity that they
were not useful for decision making. For him it was clear that the purpose
was more important than the means; that numbers, when they do not fulfill the
purpose of providing information, are useless.

One may regard “defensibility” as a weak standard for a decision on a mat-
ter as consequential as automobile emissions. But it is probably the strictest
standard we can generally satisfy with real-world problems of this complexity. Even in situations of this kind (perhaps it would be better to say “especially in situations of this kind”) an appropriate representation of the problem may be essential to organizing efforts toward solution and to achieving some kind of clarity about how proposed solutions are to be judged. Numbers are not the name of this game but rather representational structures that permit functional reasoning, however qualitative it may be. (Simon 1996, p. 146)

Consequently, the identification of how to design artifacts for use in dynamic environments is probably one case where quantitative methods are often not feasible alternatives. For this, we should rather focus on maximizing understanding, which presupposes theories, communication, evaluation and a clear purpose. The first is developed within the academic discipline of HCI, the second and third must be part of practicing HCI, and the last can only be given by a design philosophy. In previous chapters of this thesis, I have laid the foundation for one design philosophy; there may be others. In this chapter I have described how it applies to practice.
In this chapter, I give two simple examples of Theories of Use. They are intended to give an idea of how to make use of the Theory of Use protocol, but should be considered with caution. They neither show the evolution of the Theory of Use, nor are they directed at a certain target audience. I chose to include them because they are addressing problems that subsequently have been proven interesting to address also commercially.

9.1 Typing on a Touch-Screen Tablet

Typing on a touch-screen tablet is rather slow and it is easy to make mistakes. Considering that the use of tablet computers is increasing, this is an important problem to address. Can typing be improved?

9.1.1 Model

An analysis of an ordinary keyboard can perhaps give clues for how a touch-screen keyboard can be improved. The assumptions below are based on theory and observations of both regular and touch-screen keyboard users.

1. It is slower to type on a touch screen keyboard than it is on a conventional one, since the typist does not get immediate tactile verification that she has hit the right key to produce the desired letter. The typist has to constantly pay attention to the keyboard and spend much cognitive effort on where she is putting her fingers, which makes it difficult to automate typing.
2. If the typist does not have to think about where she is putting her fingers, i.e., can automatize typing, she could more confidently focus on the contents of the text rather than on the generation of letters.
3. Multiple input points imply faster typing. Even though people historically have learned to write very fast by hand, the decoding problems resulting from very fast handwriting makes it unsuitable as an input channel for computers. Furthermore, expert typists still type faster on a keyboard than is possible to write by hand.
4. Fitts’ law, as it is established within HCI, states that the time to move something to a target area is a function of the distance to and the size of the target. Thus, minimizing movement, without adding ambiguity about the target point, increases speed. If it would be possible to reliably distinguish
minimal movements of fingers, that would be optimal for an input device if it would, at the same time, have a reasonably low learning threshold.

5. People would learn to touch type if they saw a benefit in it.

6. Few people will make the effort to learn a completely new keyboard layout. As a side note, it should be mentioned that the popular conception that the QWERTY keyboard was designed to make typing as slow as possible is not correct. The design was chosen to reduce the likelihood of jamming type bars in mechanical keyboards, by an arrangement so that “the keys most likely to be struck in close succession were approaching the type point from opposite sides of the machine” (Liebowitz and Margolis 1990, p. 6). It was thus chosen to maximize typing speed on a machine that had mechanical constraints. In the end, if the keyboard layout would have been seriously dysfunctional, it would have been replaced as soon as technology allowed it.

9.1.2 Qualities

Assuming that the model is accepted by the intended users, we can identify the following qualities of a good touch-screen keyboard:

1. It increases typing speed.
2. It makes it possible to notice immediately if an error has been made.
3. It does not require to learn a new keyboard layout.
4. It makes use of multiple input points.
5. Typing can be automatized.
6. There is little ambiguity about what letter has been typed.
7. The distance to move fingers is minimized.

9.1.3 Null Hypothesis

There exist software keyboards that, compared to the regular hunt and peck keyboards, improve the conditions to some extent. Most of these make use of word prediction, based on analysis of the written content. However, they are still much slower to type on than an ordinary keyboard. The subset of keyboards that use gestures to form words also rely on moving one finger over the keys and, while that is faster than tapping with one finger, it can never become as fast as typing on an ordinary keyboard.

These gesture keyboards fulfill only a few of the qualities above. Typing speed is increased, due to the prediction algorithms, but errors cannot be noticed immediately, as the typist needs to look at the keyboard rather than at the text; one single input point is used; automatizing can be achieved on the level of letter placement on the keyboard but not for gestures for words; the ambiguity about what letter has been typed is resolved by prediction systems,
9.1.4 Hypothesis 1

By using multiple input points, typing speed is likely to increase. If we start from these qualities, it leads to the problem for the typist of knowing where to place fingers, as touch screens do not give tactile feedback in the same way as ordinary keyboards, and letters are hidden from view by the fingers. Moreover, the typist should not have to pay attention to where a specific letter is—she should focus on the text that is produced.

One way of solving this problem is to use the assumption that the user knows—or is willing to learn—touch typing. Then the home row of the keyboard can be placed where the corresponding four fingers on each hand are, rather than vice versa. The placement could be issued by a command such as all four fingers touching the surface. Left and right hand do not need to be aligned and are handled separately. For the key positioning to be unambiguous, the typist must be able to reset the position when she wants, by reissuing this command. Most likely, this keyboard needs to be complemented with a prediction algorithm, to aid the identification of separate keys that are placed closely.

Another issue that needs to be resolved is how to create the characters that are normally produced with the thumbs; most importantly, the space character. It would probably not be a good idea to use a greater area of the screen in order to fit the thumbs in it. One way to address this constraint is to rely on the prediction algorithm to separate words. This, however, reduces the typist’s control. Another way is to partly violate the assumption that people do not want to learn a new keyboard layout, by using the four-finger positioning command to mean space. This has the benefit that the keyboard is constantly readjusted to finger placement but also means that the typist has to automatize
a new way of producing space, and she cannot rest her fingers on the screen surface. Figure 9.1 shows a sketch of the hypothesized keyboard.\footnote{I made the skeleton for this analysis in March 2012, as I was evaluating the usefulness of the protocol for the first time. At the time, I was not aware of Liquid Keyboard, which is a similar solution that was presented in 2011 (LiquidKeyboard.org). In 2013, the company Dryft presented another similar solution (Dryft.com). As far as I know, neither of these has made it to the market yet.}

9.2 Smoke Detector

Most Swedish households are required to have a smoke detector installed. The more sensitive the smoke detector is, the better are the odds that a potentially life threatening outbreak of fire is noticed and extinguished before it becomes uncontrollable. As most people who die because of outbreaks of fire are poisoned by the smoke rather than burned, it is crucial to have an early alert. Therefore, smoke detectors are typically installed centrally in the ceiling of each apartment or each floor (!) of a villa, where they are exposed early to smoke. The placement in the ceiling makes smoke detectors difficult to reach for many people. This is usually not a great problem, since they tend to require little maintenance. However, if the household is small and the smoke detector is sensitive, it will go off also when, e.g., toasting bread or opening the oven door. People may, thus, be interested in a solution to this nuisance.\footnote{As with the keyboard example, also companies have been addressing this problem. This example was first presented to students in a course in interaction design in February 2012. The company Nest announced its fancy version of a smoke detector, with qualities similar to the hypotheses, in the fall of 2013.}

9.2.1 Model

Some of the assumptions below are common-sense, but it is good to also state the obvious in order to lower the requirements and stimulate thinking outside of the box.

1. Fire emits heat.
2. Fire emits smoke.
3. Heat and smoke travel upward unless a strong draft sucks it somewhere else.
4. Wireless sensors allow physical separation of sensors and their controllers.
5. Most times, people accidentally activate the smoke detector when they are doing something in the kitchen.
6. Smoke does not go easily through closed doors.
7. Heat is shielded by solid objects.
8. Sound is attenuated by closed doors.
9. People are not prepared to pay a lot for a smoke detector.
9.2.2 Qualities

A good smoke detector

1. detects fires fast
2. never fails at detecting fires
3. is easy to understand, also under stress
   a) gives a clear warning signal in case of potential fire
   b) is easy to turn off in case of false alarm
   c) gives an understandable warning message when non-functional
4. requires little effort to install
5. requires very little maintenance
6. is not dependent on abilities and physical location of the user
7. is reasonably cheap

9.2.3 Null Hypothesis

The null hypothesis, H0, as illustrated in Figure 9.2, is that status quo prevails. Most of the stated qualities of a good smoke detector are fulfilled. The common and recommended placement of a regular smoke detector, in the ceiling, however, makes them difficult to reach for many people, which violates quality number 6 above. This problem is the reason for making other hypotheses.

Figure 9.2: The null hypothesis.

9.2.4 Hypothesis 1

Hypothesis 1, as depicted in Figure 9.3a, is a stand-alone button that communicates wirelessly with the smoke detector. The user can place it wherever she thinks is most relevant. It attaches to flat surfaces with a suction cup.

This solution appears to satisfy all of the qualities stated above. Some cannot be depicted in the sketch, since they are dependent on the component quality of the devices.

The solution is building on the assumption that fire emits smoke and that smoke travels upward, just like regular smoke detectors.

A risk in this solution is that people may place the button next to their beds and turn the smoke detector off in case of fire before they even became conscious of the risk. Another risk is that the button is placed so that it is
constantly depressed. To counter these risks, the alarm should not be definitely shut down when the button is depressed but just snoozed, and the button should give a clear warning if it is constantly depressed. The warning could also be used to check the status of the smoke detector.

Figure 9.3: The first two hypotheses. These should now be showed to intended users to elicit reactions.

9.2.5 Hypothesis 2

Hypothesis 2, as depicted in Figure 9.3b, is based on the same technology as hypothesis 1. The difference is that now the button is an application in your mobile phone. When the smoke detector goes off, the user is presented with a snooze button, much like an ordinary alarm clock application.

9.2.6 Hypothesis 3

Hypothesis 3, as depicted in Figure 9.4, is based on detecting fire by the use of a heat camera mounted in the ceiling. However, if the fire is hidden by solid objects, it may take time to detect it, so the fire may already be out of hand when it is finally detected. The technology is also, relative to ordinary smoke detectors, very expensive and requires the ability to distinguish between “good” and “bad” sources of heat, since it should not trigger if the heating system or the stove is on. This hypothesis does not address the initial problem, described above, in its current design and should be used only for inspiration.
9.3 Empirical Points

Validating design principles is hard as simple post-hoc evaluation is methodologically unsound [10]. However, a more thorough theoretical framework or model of appropriation would be valuable to both validate these principles and suggest future directions of study. (Dix 2007, p. 30)

The present protocol is here put up tentatively, for falsification. I do not dare to claim that it is corroborated since, apart from my personal adoption, my practical experience with the Theory of Use protocol is limited to 17 cases of student group projects. These are in the context of a course in interaction design that I have been teaching for two years. The total of 17 projects divides into ten in the first year and seven in the second. The number of members in each group was four in 15 cases and three in the remaining two.

The focus of the projects was, in both instances of the course, to identify application areas for the Internet of Things, i.e., to imagine how interconnected sensors, actuators, and computing devices in our surroundings could become useful for people. Apart from written instructions and examples, including the ones in Chapter 9, the students were given lectures in interaction and user experience design, with special attention given to the autonomy of the user.

One may ask what a falsification of the protocol would appear like, when there is little opportunity to do explicit manipulation of variables. It is a legitimate question and it is, indeed, the case that all student groups followed approximately the same protocol. Consequently, no comparison has been made with groups that have not used the protocol, which makes it impossible to determine its value statistically. However, a qualitative assessment of the outcome can be made of the involved steps: the initial Theory of Use can be evaluated—not based on how correct it is, but on how well it communicates

---

For an introduction, have a look at the very accessible pamphlet Inspiring the Internet of Things (Presser 2012), which was also given to the students.
the thinking of the designers—and the hypotheses can be evaluated based on how well they follow the Theory of Use. This was done in peer reviews, i.e., part of the protocol. After hypotheses had been generated, the students were given the task of scrutinizing each others’ reports. The questions to be answered were:

1. Are the qualities, the situation and the model reasonable?
2. Is there something essential that has been neglected in the theorizing part?
3. Do the suggested designs, i.e., hypotheses, satisfy the theorizing that the group presents?
4. Do you understand what the group is envisioning? How could the presentation of the idea be improved?

The feedback that was generated from this falsification was considered as a very valuable contribution to the project work, and functioned for the students as a way of adjusting the assumptions that they had made about potential users, as well as for refining the hypotheses about qualities in a good solution.

The course has been well liked and the project has been considered as very rewarding. In the first course evaluation, one student reported: “I got a new way to think!” and, when I met one of the students approximately one year after the course had finished, he told me, rather excited, that the protocol had been very valuable in his continuous studies. The innovative results of the projects indicate that the protocol can support creative, yet structured thinking.

The results from these cases of application of the protocol should be interpreted with caution, as they do not constitute a controlled study. The outcome depends on many factors, of which the project structure is only one. Rather, the evaluation of the protocol follows the general philosophy that I advocate here, that of subjecting hypotheses for falsification in practice. Even if this cannot prove the value of the protocol, neither have I been able to falsify it yet.

The principle of autonomy, on the other hand, has been extensively proven in practice, although not explicitly identified as such. For instance, practically oriented projects, involving members from my research group, have for almost two decades delivered tools that follow the so-called GMOC framework for analysis: Goals, Mental models, Observability and Controllability (Andersson et al. 1997). The essence of this framework is to enable a worker to decide autonomously: it is important that it is explicit what her goals, or purposes, are; that she can observe the state of the tool that she is using so that she can form a proper mental model of her work; and that she has sufficient control over the tool.

I have stressed these empirical points, not because I wish to use them as an argument, but merely in order to avoid objections. (Popper 1966, p. 122)
In this final example, I will go back to how all of this started. This is a report of the project that I was working on during the first half of my PhD studies, and the very reason for devising the Theory of Use Protocol. I was constructing tools with the intention to stimulate ethical analysis of difficult, uncertain situations. Based on the tests that I ran on students, I gradually came to suspect that, although laudable, the effort that my tools demanded from users, without providing certitude about the moral correctness of a conclusion, made the tools practically useless. What the students wrote in their analyses indicated that most of them were just performing a prescribed ritual, without realizing any value in what they were doing. What I concluded was that a tool cannot help someone who is not already experiencing an urgent need for reflection over ethical decisions.

I have gotten a few reports from former students who have continued to use the tools for evaluating their options in different quandaries, which is very positive. However, these people probably do not even need the tool; it just makes the process more convenient. The ambitious goal of supporting those who are not used to thinking about the moral dimension can thus be concluded to have largely failed.

This does not seem unique to the approach that I chose. Judging from the adoption of other ethical decision support tools, which is limited, to say the least, moral questions are neither easy, nor very attractive, to analyze.

The Theory of Use Protocol is a result from this realization. It seemed to me that if ethics was ever to become an integral part of decision making, it needs to fill a formative function; it needs to lead to decisions that are better not only from a moral point of view. I am, however, getting ahead of myself now. Let us start with revisiting one of the scenarios that I presented in the introduction to this thesis.

10.1 The Ultimate Decision Support, Revisited

What if computers can help you make perfect decisions every time? How do you think that would affect your ability to make decisions?

\(^1\)This chapter builds on work I have done in my research group (Kavathatzopoulos, Laaksoharju, and Rick 2007; Laaksoharju and Kavathatzopoulos 2008; Laaksoharju and Kavathatzopoulos 2009; Laaksoharju 2010; Prax and Laaksoharju 2012; Laaksoharju and Kavathatzopoulos 2013).
This is a question that has no definite answer, as we do not yet find ourselves in such a situation. In the introduction I suggested that too much support in decision-making reduces our ability to think autonomously when we should. However, our decision making could still benefit greatly from properly designed support tools, and surely we need to make decisions all the time. What is important here, is to understand and draw the line between what is mere support and what poses constraints for autonomy. For this, we need to understand human decision making, and we also need to take a philosophical position about what is a meaningful life. These matters have been addressed in this thesis.

For an engineering mind, a philosophical position may seem unnecessarily convoluted and if we regard the near future—and consider everything else that we have to do—it may very well be so. However, what I am aiming for here is not just a sufficient decision-support tool; I am aiming for a morally and theoretically justified one.

### 10.2 Ethical Decision Making

The word *ethics* can be understood in several ways. I understand it as the study and prescription of morality. When I use the word *ethical* to describe decision making I, consequently, refer to decisions for which it is to be determined what is moral.

Applied ethics is not a matter that is of interest only to philosophers. Most of us have been upset by at least some examples of misconduct in political or financial affairs. Ethical aspects impact most aspects of both professional and casual life. Depending on how a decision is made, or a situation is handled, different possibilities of gains and losses present themselves, in economical terms as well as for other values like justice, dignity, and privacy, to name a few.

When it comes to ethics for corporations, there are several different positions, of which I will account for two.

Keith Davis (1960) makes a strong case for corporate social responsibility by arguing that, if corporations do not voluntarily take responsibility, they will lose the society’s trust in their ability to take it. The effect is increased regulation and decreased autonomy. His suggestions for how corporations take responsibility are remarkably in line with what is proposed in this thesis:

> making work meaningful, developing persons to their fullest potential, preservation of creativity and freedom, and fulfillment of human dignity. (Davis 1960, p. 76)

An indifferent observer, like M. Friedman (2007), does not endorse social responsibility as an obligation, but rather as a way of doing business successfully: when evaluating the morality of a certain decision, it also means that
the economic sustainability of that decision is evaluated. It is, bluntly put, bad business to do something that people in general think is unethical. However, this presupposes that consumers have the possibility to act according to their moral convictions, which may be expecting too much. We can observe that, for instance, smartphone manufacturing malpractices do not seem to affect sales of the smartphones, nor the reputation of the smartphone companies. To make moral decisions, it is important to have reasonable alternatives, and access to all relevant information.

The questions that follow are: what is relevant information and how we can determine that we have all we need to make a moral decision?

In the late nineties, Walter Maner did a survey of sixty methods, designed to investigate ethical questions in a systematic, procedural way. This survey is presented in the paper “Heuristic Methods for Computer Ethics” (Maner 2002). In it, twelve out of the sixty methods are singled out to represent different approaches. It should be stressed already here that most of these methods are not computerized. Some of the approaches are more detailed than others but what most have in common is that they rely either on normative ethics or on moral intuition.

A valuable contribution of the paper is the attempt to synthesize these different strategies into a twelve-stage process (Table 10.1) that can be used as a starting point when determining the scope of an ethical decision-making method. In his analysis, he concludes that almost every, at the time existing, method fails, or avoids, to account for all twelve stages. It is, however, not in itself a sign of bad design, as the stages are not to be seen as strict requirements for decision making procedures, but more as reminders to focus attention on possible neglects. A method that would include all of the stages would probably end up far too tedious to be useful, but the method developer can benefit from explicitly accounting for the reasons why certain stages are chosen while others are not.

It can be questioned whether stratification in stages is necessarily the best approach for achieving an optimal result. I have therefore below reinterpreted the stages as states in an iterative process, in which the different foci are still important but not necessarily the order in which they are evaluated. In such an iterative process, the analysis evolves more organically than in a process that runs in a waterfall-like manner toward a conclusion. This is also more in line with my theoretical position, as it does not imply guiding the thinking of the user as rigidly.

10.2.1 Ethical Competence

Lawrence Kohlberg (1985) was a pioneer in the classification of ethical competence. He devised tests, in which the respondents were to give explanations for why they considered one option better than another in a situation with eth-
<table>
<thead>
<tr>
<th>Stage</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Preparing</td>
<td>Cultivation of moral awareness and acquiring of ethical competence</td>
</tr>
<tr>
<td>2</td>
<td>Inspecting</td>
<td>Definition of the problem, gathering of facts, identification of stakeholders, relationships, etc.</td>
</tr>
<tr>
<td>3</td>
<td>Elucidating</td>
<td>Identification and development of facts and presumptions, identification of key issues and framing of problem</td>
</tr>
<tr>
<td>4</td>
<td>Ascribing</td>
<td>Specification of the values, interests, principles, biases etc., that are the driving forces for a possible conflict</td>
</tr>
<tr>
<td>5</td>
<td>Optioning</td>
<td>Brainstorming to develop possible alternative actions for participants to solve the problem followed by a screening to eliminate impossible options</td>
</tr>
<tr>
<td>6</td>
<td>Predicting</td>
<td>Prediction of potential consequences from the different actions identified at the previous stage</td>
</tr>
<tr>
<td>7</td>
<td>Focusing</td>
<td>Choosing of a set of stakeholders, values or issues to consider more in detail in order to identify the core ethical issue</td>
</tr>
<tr>
<td>8</td>
<td>Calculating</td>
<td>Quantification and weighing of risks, costs, likelihoods, etc.</td>
</tr>
<tr>
<td>9</td>
<td>Applying</td>
<td>Deliberation over the gathered information, possibly application of theories and/or weighing of values and arguments for and against options</td>
</tr>
<tr>
<td>10</td>
<td>Selecting</td>
<td>The choice and common-sense verification of an option possibly leading to a reiteration</td>
</tr>
<tr>
<td>11</td>
<td>Acting</td>
<td>Planning and carrying through with the decision, development of indicators to assess the consequences of the decision</td>
</tr>
<tr>
<td>12</td>
<td>Reflecting</td>
<td>Monitoring the implementation of the decision and learning from errors if any, possibly formulating a policy</td>
</tr>
</tbody>
</table>

Table 10.1: Possible stages in ethical decision making tools, according to Walter Maner (Maner 2002).
ical implications. Through judging and classifying their answers, he defined evolutionary stages of moral development. The advantage in this method is that it focuses on the argumentation for a certain choice, more than the righteousness of it. The one crucial disadvantage is, however, that he relied on certain principles to evaluate the respondents’ explanations. Consequently, his own opinions about right and wrong were affecting his assessment of the answers. To avoid this type of subjectivity, it is desirable to refrain from assessing the normative content when evaluating ethical competence. Instead, the definition should focus on the process of ethical decision making; the way that conclusions are reached (Kavathatzopoulos 1991).

This is, however, difficult. It is very tempting to compare and relate to an ideal behavior—a standardized code of conduct or a philosophy of morality—when determining what an individual’s ethical competence is. Nonetheless, this raises a lot more questions than it solves, of which perhaps the most important is: how to determine what is good behavior? If rules are followed, everything is fine and only when they are broken would it be necessary to make a judgment.

Ethical competence would thus become an expression that is only relevant when there is a lack of it. This is an approach that elevates heteronomous, responsibility-avoiding behavior to an ideal. Following the argumentation of Kant, the actual assessment of what is moral is dependent on the reasoning of the individuals that are involved in the moral problem, and should depend on what decisions are possible to make. It is, consequently, neither the behavior, nor the decision itself, that determines whether a decision is moral, but the way in which arguments for different alternative decisions were evaluated.

The common conception of a moral problem is one where two principles are heads up with equal strength. Most artificial examples that are presented in literature on ethical or moral decision making are dilemmas like this. In practice, the validity in such a conception can be questioned. Most problems with apparent moral conflicts are more likely results from lack of information: the decision maker does not have access to the necessary information to reach the best conclusion. A decision might publicly be considered as immoral, but the effects from making the opposite decision can still be far worse. The decision maker needs to be able to elaborate on the details of a problem and argue for the soundness of a decision. As Boulding (1966) writes:

Improvements in information processing, therefore, have profound ethical significance, because they remove obstacles to that widening of agendas which is one of the major components of most ethical systems. Preaching, which has been one of the main technologies of ethics, never seems to have been very effective, beyond a certain point, and it may be that the horizons of the power of ethical ideas may be substantially extended by the development of improved methods of information processing by the individual and by the organization. (Boulding 1966, p. 167)
As we today have the technical possibilities to support information gathering, it becomes very attractive to explore how to make best use of these. All relevant information should be gathered and all stakeholders and their interests taken into account in order to reach a state when a decision can be considered to be well-founded. This motivates a focus on autonomy, as defined in Section 6.3. Autonomy is the necessary disposition to achieve true ethical competence: not to always act according to rules or guidelines; not to act in a manner that is consistent with the highest number of normative philosophies; but to maintain focus on identifying the relevant aspects of moral problems.

10.2.2 Related Work

Many of existing tools to address ethical concerns make either an implicit or explicit assumption that using ethical theories can help us make more moral decisions. Consequently, a common way of reasoning about ethics in artifact and tool development is to focus on identifying obligations and responsibilities of stakeholders, especially of those who are procuring, designing, developing, and monitoring the artifacts (Collins, Miller, et al. 1994; Mancherjee and Sodan 2004). This, however, presupposes that what is morally correct according to normative ethics is readily identifiable, which may be an overly optimistic assumption, given that most of these ethics were detailed before the introduction of computerized artifacts. Different ethical decision-support tools have targeted this problem in different ways.

In Paramedic Ethics (Collins and Miller 1992), the process starts with listing all alternative decisions that the analyst can imagine. Stakeholders that are directly affected by the different decisions are identified. The analyst then establishes all relationships between stakeholders, in terms of obligations and responsibilities. The second phase of the process is to analyze how the alternative decisions affect the stakeholders, followed by the third phase, in which the objective is to determine whether it is possible to find compromise solutions. The ambition is to identify a negotiated social contract solution that could lead to a consensus decision. The fourth and final phase is to evaluate the different alternatives in light of deontological and utilitarian ethics, to determine if it is in accordance to these two types of normative ethics.

SoDIS Project Auditor (D. W. Gotterbarn 2002; D. Gotterbarn and Clear 2004; Gotterbam 2007) is focused on identifying and mitigating risks in different tasks. It involves a computerized implementation of the Paramedic Ethics protocol, expanded with various functions to assess risks in different decisions. The user is first gathering extensive background information about the problem and its stakeholders and is then prompted to answer questions aimed at identifying known causes for moral problems.

These rules include: Don’t kill, Don’t cause pain, Don’t disable, Don’t deprive of freedom, Don’t deprive of pleasure, Don’t deceive, Don’t cheat. Keep your promises, Obey the law, and Do your duty. [...] Gert’s ethical principles have been combined with ethical imperatives from several computing codes of ethics to reflect the professional positive responsibility of software developers. These principles have been framed as a set of 32 questions related to stakeholders in a software project, and to generalized responsibility as a software professional. (D. W. Gotterbarn 2002, pp. 159–162, Stress added)

The risk in this approach is that it seems to imply that what is ethical is to answer detailed questions, without necessarily involving reflection. It details what to do and not do, but not why this should be done. The most important aspect of the codes of ethics that contribute to the analysis tool may thus be forgotten. In the ACM Code of Ethics, it is formulated as follows:

Questions related to ethical conflicts can best be answered by thoughtful consideration of fundamental principles, rather than reliance on detailed regulations. (ACM 1992a)

In ETHOS (Mancherjee and Sodan 2004), the user is advocated to identify the open moral questions at hand, through taking the role of a moral agent, after which the utility of alternative solutions are quantified according to ethical theories. If the theories seem to suggest conflicting decisions, the user is urged to find which one is more applicable to the situation.

It should be noted that the first two of these tools are intended for computer professionals working in technical development projects, while ETHOS is not targeting a specific audience and does not assume any specific content in the problem to be analyzed. Nevertheless, the approach is normative and makes references to ethical theory.

This work is based on normative, western ideas about what is good and just and these tools are, consequently, designed to work as pilots for the problem owner. None of these has had significant impact outside of their respective research groups and I think this is largely due to the fact that they reinforce obligations rather than facilitate deeper understanding of the analyzed problem. They are based on the assumptions that they should be possible to use without being knowledgeable about ethical theories, and without necessarily having a guiding philosophy.

Here it should be remarked that my position is that these tools build on a misconception of the intentions of the philosophers behind the implemented philosophies. Kant, for instance, is often interpreted as advocating duty (deontology), but I believe this to be a confusion about what this duty was directed toward (see Subsection 7.3.2). John Stuart Mill, who is the founder of utilitarianism, saw justice and maximizing of happiness as means to protect the autonomy of individuals.2 Any implementation of these philosophies should,

---

2“The moral rules which forbid mankind to hurt one another (in which we must never forget to include wrongful interference with each other’s freedom) are more vital to human well-being
consequently, focus on allowing individuals to be autonomous. Guidelines, rules, systems, and tools that reduce the possibility to reason autonomously are thus inconsistent with that purpose.

All above methods are probably excellent for systematizing, organizing, and guiding the user in moral appraisal, but there is a risk that the user’s focus is diverged from the obvious main purpose, namely to handle the practical problem in a satisfactory way. Of course the latter is the espoused purpose of all methods, but as they include comparison to different normative ethics, and in some other cases even propose moral solutions (for example Davidrajuh 2008), they shift the attention from the urgent to the artificial. They do not focus on, nor address, autonomy, which means that they cannot guarantee that decisions are not founded solely on heuristics, biases and authority obedient dispositions, i.e., heteronomy.

10.3 Qualities of a Tool for Ethical Deliberation

It goes without saying that it is very hard to decide about moral problems. It is, in a sense, the definition of the topic. Questions about morality are often debated and regard the core of human existence. The conflicting principles and values behind stakeholder interests will all seem too important to trade off, and the outcome of any realistic decision will appear to have undesired features. This is the common perception of an ethical problem.

However, in real life, many moral problems occur instead from the lack of information or misinterpretation of responsibilities. The intentions might be good, but if important information is missing, the wrong decision may be taken.

The purpose of a support tool for ethical deliberation is not to replace, but to complement, codes of conduct and guidelines. It should help decision makers in the process of establishing a conception, as complete as possible, about a problem with moral implications. The most apparent area of use for this kind of tool is in public policy making, where stakeholders are numerous and the interests of these play a key role for what decision is essentially made. It is, however, not the only imaginable application area. Moral problems are generally complex, and sometimes involve a large number of stakeholders and interests.

It is therefore important to allow the person who is facing the problem to freely add information to the analysis whenever there seems to be a reason for it. To make a well-founded decision, it is desirable to collect as much data as possible. The problem with massive amounts data is, however, apparent: the chance to make use of it decreases with the amount.

\[\text{than any maxims, however important, which only point out the best mode of managing some department of human affairs.” (Mill 1863, p. 87)}\]
Many of the approaches presented in earlier work suggest different strategies to eliminate aspects that are not relevant for the problem (see stage 7 in Table 10.1) so that focus can be maintained on the most critical aspects. The impending risk with elimination the loss of important contextual understanding. A better approach could be to allow the decision maker to be selective when it comes to analyzing the data.

The qualities here build on the assumption that ethical competence is equivalent to a well-functioning problem-solving strategy (Kavathatzopoulos 2003; Kavathatzopoulos 2004; Erlandsson and Kavathatzopoulos 2005). It targets a wide audience, and does not assume any specific content in the problem to be analyzed, which makes a piloting approach—in which the user is prompted with questions based on known reasons for moral problems—impossible. Also, since what is a moral decision—if the outcome is assessed—cannot be assumed to be determined a priori, this kind of tool must be designed so that it does not make any assertions of the normative correctness of any decisions or arguments.

This absence of guidelines may initially seem like a weak strategy, but it is a strength when it comes to widening the agenda for the problem situation. The user is never lured into the false comfort of believing that the analysis is finished. Such a setup puts the responsibility for a satisfactory analysis completely on the user.

To create the best possible conditions for a well-founded decision, it is desirable to allow the user to input any imaginable information, without having to worry about the extent of data muddling the clarity when time comes to make a decision. It is often stated that brainstorming is a good technique to generate ideas for problem solving, and in such a procedure, all ideas are initially accepted, even the obviously unsuitable ones. The purpose is of course to encourage and stimulate associations.

This idea generation process can easily be inhibited by cumbersome processes or premature overanalyzing. The inputting of information should, thus, be simplified as much as possible, so that any piece of information that might affect the decision is added. The user should not have to wonder where to add it.

Any relationship between stakeholders should be evaluated and also tried on other pairs of stakeholders. What is valid for one pair might also be valid for other pairs. The uniqueness of a particular relationship must be carefully questioned and if it remains unique after scrutiny, the reasons for the special connection should be noted.

To support autonomy, the main requirement on the tool is that it should not be making any decisions and not even directions about the correctness of any specific conclusions: it should not be elevated to an authority. The sole intention should be to help the user to organize and structure a problem. At the same time, the problem should not be narrowed down, by that risking
oversimplification, but rather widened so that the user can appreciate the full impact of a decision.

10.3.1 Theory of Use
The identified assumptions and qualities of a good tool for supporting ethical decision making are stated below.

**Assumptions about the user**
1. Does not necessarily know the different ethical theories
2. Is not necessarily trained in causal and consequential reasoning

**Assumptions about the stakeholders that are taken into account**
3. Do not necessarily share ethical principles, codes, laws or policies
4. Do not necessarily share moral values

**Qualities in a good tool**
5. Does not require or derive from a predefined set of moral principles and values
6. Helps the user to systematically solve the moral problem at hand
7. Helps the user to be unconstrained by moral fixations and authorities
8. Helps to identify and consider as many relevant values and alternate actions as possible
9. Encourages the user to give motivations for his or her decisions in regard to relevant interests and values
10. Helps the user to organize and analyze the facts
11. Helps the user to weigh the relevant values and principles against each other
12. Blocks heteronomy and supports autonomy,
13. Organizes interrelationships and data in a systematic way
14. Presents the complexity of the issue in a comprehensive way

In the following paragraphs I will describe how these qualities were expected to be fulfilled in the tools. The items that regard shared values and ethical theories are handled implicitly by deploying a non-normative approach. The remaining are handled by supplying a structure to support the understanding of how information about moral problem relate to the interests of the involved stakeholders.

10.4 Autonomy-Oriented Analysis
The approach in my work is stakeholder-centric in two ways. 1) The objective is to facilitate dialogue between stakeholders, rather than giving directions
based on the thinking of moral authority figures. 2) The analysis starts from identifying the interests of the individuals, groups, and organization that affect, or are affected, by a decision.

I have not attempted to create a tools that tell the user what is objectively right. On the contrary, they have been designed to create uncertainty about dogmas and what to consider as objective, universal truths. Since the early 90’s, my supervisor Iordanis Kavathatzopoulos has used this approach to train decision makers in different contexts, notably in business and politics (Kavathatzopoulos 1993; Kavathatzopoulos 1994; Kavathatzopoulos and Rigas 1998; Kavathatzopoulos 2003; Kavathatzopoulos 2004; Kavathatzopoulos and Rigas 2006). The assumption behind this objective is that in order to understand the points of view of others it is necessary to understand the foundation for one’s own convictions. Thus, these need to be equally scrutinized.

10.4.1 Support Self-Scrutiny

The process of self-scrutiny is, however, difficult. Sometimes conclusions are uncomfortable or even distressing. In such cases people may prefer to ignore or suppress information that is pointing in a direction that does not feel right. Medieval astronomers’ efforts to rationalize the geocentric world view is but one example of this (Feyerabend 1993). On the other end of the comfort spectrum we find information that is pointing in the direction of accepted truths. Confirmation bias makes it easier to accept information that supports a claim rather than undermines it (Sunstein 2005). This type of information has deserved the special name “evidence”. The software that I develop is not designed to formulate evidence but rather to map arguments for and against different decisions. In addition to being stakeholder-centric, it is also argument-centric. More precisely, arguments are founded in the stakeholders’ values and interests.

Our values are the fundament of everything. This has the curious effect that they are also used as temporary solutions to difficult problems. Throughout history, societies have created taboos and protected values around topics that are hard to argue consistently for or against (Baron and Spranca 1997). Even what is considered as logical can be made to adhere to values, which implies that it can sometimes be disturbing to scrutinize the logic of some sensitive issue. For instance, in 2012 two Italian philosophers took the idea that newborns do not gain the self-awareness that would qualify them as individuals immediately after birth to its extreme, by making a case for after-birth abortion (Giubilini and Minerva 2013). The article generated immense criticism and hatred against the authors and is a vivid example of how we become emotionally engaged when our values are threatened, at the same time as the criticism shows that we tend to accept arguments less critically when they conform to our values.
However, not only do we have problems with arguments that may lead to inconvenient conclusions, we also have problems understanding arguments for unfamiliar positions. For arguments to be accepted, they need to be presented in a language that is understandable for the audience. This has the uncomfortable side effect that acceptable arguments could be used to argue for something that is not considered acceptable. One classic example: if both logical consistency and tolerance are highly revered in a society, a paradox is created where one may get the impression that consistency requires that intolerance should also be met with tolerance. The philosophers John Rawls (1999) and Karl Popper (1966) solve this dilemma by posing the constraint that tolerance only applies as long as the conditions for it are not threatened. To tolerate intolerance would lead to the end of tolerance and is, consequently, not included in what should be tolerated. Most of us probably accept this conclusion but if we deconstruct it, we may find that it is rather arbitrary and founded more in values than in logic. How do we for instance classify what is intolerant? This kind of doubt is important in order to develop one’s own thinking. The purpose is not to encourage nihilism, but to understand the underlying values behind strongly held opinions, and through that gain humility for the possibly conflicting opinions of other, equally self-reflective, individuals.

10.4.2 Bias Countering

To support the decision-making process, a support tool would optimally either block biases or make these apparent to the user. The first and most important is the bias to regard only the information that is available, which is then evaluated to support the current position. By encouraging the user to expand the problem scope, and by presenting information in a clear structure, this may be countered. With an expanded view of the problem, the risk for underestimation and oversimplification will also, most likely, be reduced.

By a systematic scrutiny of stakeholder interests, which leaves the determination of alternative decisions to the last steps, the risk for unconsciously omitting relevant information decreases, as most of the work for analyzing more than one option is already done. There will be less incentive to terminate prematurely or leave out information, which are the observable effects of choice-supportive and confirmation biases. If it is possible to easily create compromise decisions, it could help counter the fixation on certain alternatives.

Biases toward specific stakeholders, like in-group favorability bias and the fundamental attribution error can be countered by first letting the user take the perspective of all stakeholders, and subsequently depersonalize these. When analyzing arguments, the identities of stakeholders can be obscured, which will leave only the statements as basis for a decision.
10.4.3 Support Dialogue

A dialogue between different cultures, for instance between socialists and libertarians, will rarely lead to mutual understanding because the sets of values and the apperceptions of the world differ too much. They will simply not interpret the opposing faction’s arguments as they were intended. This phenomenon is something that I try to take into account when designing the software. In order to facilitate understanding between differing cultures or ontologies, it is focused on the fundamentals: the involved values. The point is to elucidate assumptions that otherwise are left tacit; assumptions about what matters in a situation. These assumptions determine the interpretation of a situation and a hypothesis is that if arguments are presented in a consistent form, focused on how decisions impact stakeholder values, that will curb rhetorics and facilitate understanding, which opens the possibility for dialogue.

10.5 Support Deliberation

A tool should refrain from dictating what is right and wrong, since that would necessitate a fixed moral standard. Rather it should support a philosophical inquiry and let the moral implications be determined in the process of analysis, as this is dependent on what values are considered. This may seem like a risky approach, as it gives no guarantees that everyone will agree on the morality of some decision.

However, it is, as Kant argued, not enough that a conclusion happens to be in accordance with some moral rule in order to be truly moral. Even a computer, which cannot (yet) be regarded as a moral agent, can follow rules blindly. Furthermore, we expect societies to comprise mechanisms to define and communicate moral values between its members. A tool should, consequently, be formative, i.e., suggest what aspects to consider and how to formulate arguments that are suitable for dialogue, but not normative by determining what conclusions are correct.

Through association and invitations to fill out blanks, further considerations are triggered. Collins and Miller (1992) agree with the benefits from expanding the problem through an associative process, and their Paramedic Ethics is in many aspects a logical effect of this assumption. There is, however, one important difference: in Paramedic Ethics, stress is on normative obligations, vulnerabilities and opportunities, while what is focused on here is how the interests of different stakeholders relate. This makes the analysis procedures different. Where Collins and Miller urge normative evaluation, this process only creates a foundation for thorough, autonomous analysis. The assumption for the present approach is that an unconstrained and non-moralizing approach liberates the decision maker and enables her to take into consideration aspects that she considers to be relevant.
10.5.1 Support Construction of Arguments

In an introductory textbook to moral philosophy, James and Stuart Rachels write that it “would be nice if there were a simple recipe for constructing good arguments and avoiding bad ones. Unfortunately there is not” (J. Rachels and S. Rachels 2012, p. 12).

Despite this claim, I will suggest a recipe, with the sole intention of making arguments that are suitable for dialogue. The goodness of the arguments is, thus, in regard to its form and not its content. Such an argument specifies who a certain measure concerns, why it is relevant for her, and how it affects her.

Furthermore, it is explicit about what the risks are in some decision, at the same time as it presents the possibilities. This can be achieved through the form of the analysis. Below is an example of an argument that is structured to open dialogue. The stressed words correspond to the above criteria.

The possibility to participate anonymously in public debate is important for citizens because they can then voice concerns without fear of retribution. There is a risk that such channels will be used to spread anti-democratic ideas, but there is also a possibility that some specific concerns are widely shared and thus necessary to bring to attention. ICT tools can be constructed to stimulate users to engage in constructive dialogue rather than one-directional polemics. There is a risk that such tools will limit the people’s possibilities of expression so great discretion is necessary.

10.6 Hypothesis Testing

So how can a Theory of Use for an ethical decision-making tool be tested?

It would make little sense to set up a controlled study, in which I compare the tool with a tool from the piloting approach. Even if I would be able to gather test subjects that are willing to perform several analyses over a couple of hours, it would be hard to operationalize problem understanding in a way that would be agreed upon by both camps.

While both approaches focus on stakeholders, the piloting approach tries to create an understanding of which stakeholders are impacted by a solution, how they are impacted, and what their responsibilities are. In contrast, the approach that I have worked from aims to identify the relevant values of stakeholders and, based on these, what the arguments for and against certain decisions are.

The outcomes from the two types of tools would be difficult to compare, since one gives guidance to make a normatively acceptable decision, while the other only yields uncertainty, possibilities, and risks. The same problem applies to a qualitative comparison. The results will be interpreted differently depending on what is considered important.

An attractive way of studying this kind of tool is by regarding it as a hypothesis in itself, which can be falsified in actual use: the test subjects use them for
the purpose that they were designed for and it is determined from their success or failure to make good decisions whether they work or not.

However, this again poses a severe problem: how do you know what is a good decision? For that we need to make a philosophical assessment, since neither would it be enough that arguments are accepted by a majority of the population (few would admit that Germany during WW2 made only good decisions), nor would it be enough that the outcome seems good within the observable future (DDT is an example of this type of fallacy). The assessment is, thus, difficult to make based on the outcome. Instead, we need to consider the decision-making process and assess that, based on an idea of what we should strive for as humans. The only feasible measure seems to be how well the software supports our ability to reason under uncertainty, which can be determined by the way that arguments are presented.

10.7 Iteration 1: EthXpert

EthXpert is the first tool that I designed for autonomy-oriented analysis. The purpose was to aid primarily the analyzing part of a decision process and it was intended to be used by a group of analysts working together.

The reasoning behind the tool was that an analyst needs gentle nudging in order to overcome her own biases. Based on trials with a similar pen-and-paper method (Kavathatzopoulos 2003; Kavathatzopoulos 2004; Erlandsson and Kavathatzopoulos 2005; Rauschmayer et al. 2009), I concluded that the matrix representation, with stakeholders and interests on one axis and alternative decisions on the other, serves fairly well for promoting a systematic assembly of available information about a problem. The matrix cells are filled with the possibilities and the risks in each decision alternative for each interest of each stakeholder.

The question that I approached with EthXpert is how this process can be improved by computerization. The pen-and-paper method supports a holistic overview, and a naïve spreadsheet implementation of it would lack its necessary flexibility and overview. The desired systematization would also benefit from a less limited and more associative process of inputting the data, as well as from a more configurable visualization of the same. Furthermore, the process of identifying interests and how these interests are involved in a situation would benefit from relating interests to other stakeholders; from questioning the uniqueness of relationships and from reusing information to stimulate further consideration. These features, for which computerization is necessary, were introduced with EthXpert.

From Maner’s above set of stages (Table 10.1 ) the tool covers stages 2 to 6: inspecting, elucidating, ascribing, optioning and predicting.

Stage 7: focusing, was omitted because it implies that the analysis must be narrowed down. In EthXpert the decision to focus the considerations must
be a responsibility of the user. Stage 8: calculating, was omitted because it risks fostering a heteronomous attitude in the analysis, by making isolated judgments and subsequently using a convenient figure to make a decision. Stages 1 and 9 to 12 regard the actual decision making. If the tool is used within organizations, which is the most likely case, stage 1: preparing, and 12: reflecting, involves the establishment of roles and routines to increase the ethical awareness in the organization. Stages 9 to 11: applying, selecting and acting, requires well-functioning group and managerial processes. As I am more concerned with identifying the information that is necessary for a holistic analysis, I will not discuss them explicitly here, and instead refer the curious reader to previous work in my research group (Kostrzewa, Laaksoharju, and Kavathatzopoulos 2010).

To stress that the procedure in EthXpert is not a linear series of steps, I came to adopt term state in place of stage. Three details are especially interesting to note in the representation of the ethical procedure in Figure 10.1. The first is how the flow, from defining stakeholders to defining how interests relate to other stakeholders, discourages defining options before the initial set of relations have been established.

The second is that it is possible to reach all other states when you are defining options. This is a deliberate strategy to put focus on elaborating arguments and stakeholder relations.

The final peculiarity is the lack of a final state, which implies that an optimal state can never be reached. This is in accordance with what Boulding
(1966) means when he is writing about the necessity of widening agendas. From a deterministic perspective, it is unsatisfactory, but from a self-critical perspective it is highly desirable: the user will never be seduced to believing that the analysis is completed, but must make this decision by herself.

What eventually is to be considered as a solution to the analyzed problem must be negotiated in dialogue between peers. Individually, we are constrained by our senses and experiences, which means that we can never be certain that we have found the right solution. As displayed in academic discussions, scrutiny and argumentation are at the core of establishing what is true, and this process should not be completely abandoned just because most opponents have been convinced. By this, I do not claim that it is desirable to debate every single assertion, but a healthy skepticism can guide individuals to pursue investigations with the goal of truly understanding your own reasoning.

EthXpert supplies configurable representations, where only the data associated to a specific part of the problem is viewed. For defining such a subset of data it can immediately be concluded that it is not desirable to leave the selection process to a computer. A computer could perhaps be programmed to choose arguments based on an algorithm for ethical analysis, but as Maner concludes: “Ethical problems are too complex and too fluid to solve algorithmically in human time” (Maner 2002, p. 340). Furthermore, automating the definition of what is relevant is a sure way of elevating heteronomy as the user may start to rely on the tool to make the right decisions. Instead the subsets have to be defined by humans, specifically for each problem. The only way to solve this is to take advantage of the computer’s ability to organize and visualize data, by querying for relevant information such as affected stakeholders, interests etc. The decision maker will thus be in control of choosing information, while the tool will only assist with the bookkeeping.

Before starting the procedure, the problem should be described in as much detail as possible, not leaving out any information but still without asserting conditions that are not proven. The description should be detailed enough so that any other person could understand what the problem is. The purpose with this simple task is to establish a clear focus on the problem and the factual conditions of it.

The general idea of the tool is to counter biased judgment, and suspend the tendency to involve emotional judgment and prejudice, by dividing the analysis into two main parts. In the first part the relationships between stakeholders are established, and in the second part different decisions are considered. In the latter, preference for certain stakeholders or a certain decision should not be determining how the information is evaluated. To approach this, the relationship considerations are presented in a way that anonymizes the respective stakeholders.
Define Stakeholders
It is not an obvious task to identify the stakeholders that affect or are affected by a decision. The concept of stakeholders has not always been unanimous even among scholars specialized in stakeholder theory (Donaldson and Preston 1995). The notion of stakeholders could be elaborated, but for the present application a wide definition is sufficient: A stakeholder can be any individual, group, organization, or even society, that affect, or are affected by a decision.

Everybody who affects or is affected by a decision deserves to be at least considered in the analysis. By focusing on interests of stakeholders, associations are triggered. For each stakeholder that is directly involved in the situation, there may be secondary stakeholders that could also be influential, even if indirectly. In the tool, an implicit question will guide the user to increase the problem scope: who is affected by or affecting a specific interest of a stakeholder?

A review (see, e.g., Goodpaster (1991)) suggests that very few of existing tools to address stakeholder needs seem to have the same explicit focus on how interests affect and are affected by decisions. Many tools, in fact, seem to aim at identifying and ranking the importance of stakeholders or some other factors (e.g., Moodley, Smith, and Preece (2008)).

Define Interests for Stakeholders
For each stakeholder, the interests that are relevant to the situation should be identified. This includes the interests that might have an impact on other stakeholders. It is assumed that the explicit focus on interests of the stakeholders will help the user to identify possible conflicts between stakeholders, but also to widen the scope of the problem. The implicit function of focusing on stakeholder interests, rather than stakeholders, is that it naturally brings values into the analysis. Goodpaster (1991) calls this a Multi-fiduciary Stakeholder Synthesis.

The user identifies for each stakeholder a set of relevant interests, which could very well be unique for each of them. All interests that might affect or be affected by other stakeholders are important to consider and in the process of scrutinizing interests additional stakeholders will naturally become involved in the analysis. In the further process, some of these interests might, as you would expect, be proven irrelevant, but the narrowing of the focus should be left to last.

Define how Interests Relate to Other Stakeholders
The user is encouraged to imagine how the identified interests affect and are affected by other stakeholders to form an interest-stakeholder matrix. A relationship can mean actively affecting stakeholders or passively being affected by them. It can also regard a mutual relationship. This is the core of the analysis and draws a picture of the dynamics of the ethical problem. An interest of
a corporation to maintain a steady cash flow may put pressure on the corporation’s research department to produce salable results, which in turn can create a conflict between quality and productivity at the department, etc.

This process also helps identify previously unidentified stakeholders. The topics that are brought up in one relation may raise associations to other stakeholders. Sometimes such secondary stakeholders prove to have important influence on the dynamics of the problem. Explicitly stating how the interests affect and are affected by other stakeholders gives a background for the further analysis of scenarios from different decision alternatives.

**Define Main Decision Alternatives**

After all the relationships between stakeholders have been exhaustively analyzed, the user can start considering what options there are to handle the problem. This will constitute the *interest-decision matrix*. The most apparent alternatives for handling the ethical problem can be immediately stated. These are usually mutually exclusive and similar to answering some question with “Yes” or “No”.

**Translate Considerations**

The considerations from the interest-stakeholder matrix are not automatically copied to the interest-decision matrix. Instead the input dialog summarizes all the relevant previously stated considerations in a depersonalized way. These will serve as background, stimulation and incentive for considering how the different decision alternatives affect the stakeholders. For each alternative course of action the user is urged to state how the interests of each stakeholder is affected if that would be the final decision, including both possibilities and risks.

**Define Compromise Decision Alternatives**

To counter problems in the main options, i.e. unacceptable negative effects, compromise scenarios can be spawned from existing options. A compromise option will inherit previously made considerations from the option that it was derived from, but the user should revise these to determine the difference in effect between them. This feature is useful for considering many options that only differ partly. Ideally the user easily gets an overview of the strengths and weaknesses of similar alternatives when being able to focus the comparison by concentrating on the effects that they have on stakeholder interests.

**10.7.1 Evaluation**

In 2010, a small study was performed to evaluate the tool. In this study 5 women and 6 men, of four different nationalities, between 25 and 35 years old, all of whom were PhD students at the Faculty of Science and Technology
at Uppsala University, were given the task to analyze an ethical problem that they considered to be relevant and important to themselves.

To analyze the EthXpert tool, we should revisit the qualities and assumption that were made about it. Based on the results of the qualitative evaluation, it is suggested that the tool did not provide the qualities 10 and 11. Comments from the users included (Laaksoharju 2010, pp. 60–61):

- [The analysis] turned into a combinatorial explosion, fields that needed filling in expanded rapidly
- [The analysis] grows fast and becomes impossible to handle
- [The analysis yields] too much information
- [I] lost the overview due to the high number of stakeholders and interests
- [It was ] hard to select what is important
- [It was] difficult to explain, discuss and solve the problem
- [The tool did not provide] a good overview when deciding options

The matrix that was used to structure the analysis only provided superficial structure. In practice, the over hundred cells that some of the analyses yielded were not possible to maintain an overview over. Consequently, the tool did not give the necessary help for making a final judgment. This led me to reconsider the purpose of the tool.

10.8 Iteration 2: From Analysis to Collaborative Deliberation

Societies are systems of people and thus societal problems can only be solved by addressing the values and interests of involved people; by answering questions about how these are affecting the problem situation and how these are affected by any proposed solution. Furthermore, it is desirable to allow as many of the stakeholders to contribute to the decision-making process by adding their perspectives and considerations.

To better cope with the latter quality, EthXpert evolved into a web-based platform that could facilitate concrete and constructive participatory decision making. This collaborative tool was tentatively named ColLab, as a portmanteau of Collaboration and Laboratory (Alsadat Kazemi and Eskandari 2013).

The tool supplies an open-for-all, distributed platform, in which analyses of situations can evolve organically. It allows different parties to include their
points of view, and decision makers to follow how arguments have been applied to concrete situations. The central idea is that a tool for decision making should confront the users with how their decisions influence all involved stakeholders and their respective interests (Kavathatzopoulos 2003). It requires the users to first compile information about the situation, i.e. about the stakeholders, their interests and their relations.

Commonly, conflict situations are created by stakeholders having incompatible interests. Even if personal interests are similar, there may still occur conflicts. One example of this would be an interest in performing religious practices, which may be in conflict with another religion’s rituals or values in a secular society. There can thus be tensions, rivalry, competition or any other shared conflict history influencing a situation and any one decision in a conflict means great risk of hurting the interests of a part of the stakeholders. This is what makes up ethical problems, also outside the realm of principled reasoning.

However, by viewing a possible decision from the point of view of every other group and with their, often very understandable interests in mind, the user will be supported in countering preconceptions and other cognitive biases (Laaksoharju 2010). The form in which arguments are presented makes it difficult to revert to value-laden principled reasoning, and invites to a proactive, concrete, solution-oriented discourse, which also decreases the power of rhetoric and invites pluralism. Unlike in debates, there is no need to limit the number of issues. Structure is instead given by applying arguments to concrete interests of stakeholders (see Figure 10.2).

By gaining an understanding of the social implications from one’s own decisions, the user will be in a better position to fulfill Kant’s maxim to act so that the decisions one makes can be considered right, regardless of who makes them and who is subjected to them (Kant 1785/2004). The tool thus gives support in determining the decision that after scrutiny is most consistent with one’s own morality. At the same time it is developing awareness of moral problems and the skills to deal with these. As it is based on the assumption that people are not making judgments in isolation, it allows multiple parties to contribute to an analysis by adding stakeholders and interests; give suggestions for decisions; and participating in analyzing the consequences that these may have. This means that it functions as a facilitator of democratic participatory and inclusive dialogue (Kavathatzopoulos 2010).

10.8.1 Democracy

Democracy seems to have become one of those words, along with words like liberty, privacy and justice, that in academia requires definition every time they are used, while they in much of society are understood as the formal frameworks that are created to ensure them; frameworks like separation of powers,
Figure 10.2: A sketch of the main interface for the next revision of CoLab.
protection of personal information, everybody’s right to vote and freedom of speech. One would suspect that the real understanding of these terms is either disputed or left vague on purpose, since they are used to address diverse and sometimes conflicting interests. In his 1946 essay *Politics and the English language*, Orwell voiced a similar suspicion:

> In the case of a word like democracy, not only is there no agreed definition, but the attempt to make one is resisted from all sides. It is almost universally felt that when we call a country democratic we are praising it: consequently the defenders of every kind of regime claim that it is a democracy, and fear that they might have to stop using that word if it were tied down to any one meaning. (Orwell 1946)

Concerning the image of democracy, the recent political turmoil in Europe serves to suggest that focus on formal democratic rituals, without a deeper understanding of their function, may lead to issues that are difficult for democrats to handle satisfactorily. In Italy, Beppe Grillo’s M5S movement has had electoral success in questioning politics that do not seem to care for the citizens. In Hungary, Viktor Orbán’s Fidesz party has gotten parliamentary support for changing the country’s democratic constitution, based on the argument that traditional family values have to be protected. In Greece, the authoritarian Golden Dawn is leveraged by discontent with the current system, which is perceived as corrupt.

Apparently many voters around Europe do not have the impression that existing democratic institutions serve their interests and seek for alternatives where they can find them. In other words, the impression seems to be that established politicians are not fulfilling their role as representatives of the citizens.

The intention with this selective description is not to paint a full picture of what democracy means for different people; it serves only to motivate why I in the present proposal will assume and try to handle a widespread confusion about means and ends of democracy. I claim that democratic frameworks are not ends in themselves and exist only in order to protect the democratic process, which I will detail further below.

If we would work from a pretense that democracy actually implies public participation in decision making, we might ask ourselves how this has been affected by recent developments in ICT. This is a reasonable question as we can observe the impact of ICT on various societal phenomena. I will not answer this question specifically, but rather propose a theoretical model for how to design ICT tools so that they do support democratic dialogue.

To investigate the matter of ICT in relation to democracy, I should first present a definition of the latter. I will thus defy Orwell’s observation that the term does not easily yield to one particular interpretation. Schmitter and Karl (1991) proposes a fairly non-normative definition for how citizen influence should be guaranteed:
Modern political democracy is a system of governance in which rulers are held accountable for their actions in the public realm by citizens, acting indirectly through the competition and cooperation of their elected representatives (Schmitter and Karl 1991, p. 76).

In order for this to be realized, the modern democracy has to supply “a variety of competitive processes and channels for the expression of interests and values” (Schmitter and Karl 1991, p. 78). This is in line with my understanding, that democracy is essentially a dialectic process to reach collective decisions that take into account, but not necessarily reproduce, all of the citizens’ interests. A democratic society is, thus, not one in which the majority rules, and not one in which a certain set of eternal values are protected, but one in which interests and values of citizens are negotiated and balanced in a reflective process, by representatives that can be held accountable for their decisions. Democracy is a rational negotiation process with the aim of achieving solutions to shared problems.

Unfortunately, this is not the only existing definition of democracy. There are many interpretations of what it constitutes and not all of these recognize that it presupposes an acceptance of dissenting ideas. The common-sense definition is either result-oriented, focused on formalistic aspects, or both. The first implies that societies that are providing high living standards, security, tolerance, and other goods are called democratic. The second is based on the existence of certain procedures, institutions, roles, and processes. The qualities that are commonly seen as constituent of democracy, like separation of powers; equality; elections; freedoms of speech, political expressions and press; etc., are, however, meaningless if there is no public debate (Held 2006); if only the predominant, publicly endorsed ideologies are ever voiced. The fallacy in this is that presence of formal procedures is sufficient for democracy.

If we rather regard democracy as a process, neither the result of the process, nor its superficial formal characteristics are the most significant aspects. Democracy does not necessitate specific ideals, nor a specific set of rituals. Sustaining the democratic process becomes the most important consideration, which indeed necessitates the existence of formal procedures, but only as auxiliary to it.

In essence, democracy is dialogue between people, and this dialogue is what has to be encouraged, supported and protected. Citizens need to be involved in searching for solutions to problems, by thinking together with others. This presupposes that each person has a corresponding dialogue within herself and that each person starts from the position that the own thinking can be improved, i.e., liberated from false beliefs. The latter is the greatest challenge. The mindset of each participant in an ideal democratic process, or a dialogue, needs to be such that she always feels the need for other participants, simply because she is expecting to develop better ideas together with other able people (Sen 1999).
Nevertheless, we cannot disregard that dissent can be destructive and even pose a threat to the democracy that embraces it. Recently we can observe worrying trends all over Europe: that populist parties are gaining increasing support and influence, resulting in policies that could become threats to principles like equality of individuals. Such tendencies, which are appealing to the self-interests of a voting majority, must be dealt with by stressing the only core principle of democracy, i.e., the open dialogue between all.

However, we must not lure ourselves to believe that this is an easy task. Not all are equally well predisposed to deliberating and persuading others in a way that is accepted in the common political discourse (Sanders 1997). This poses a threat to democracy, but ICT can play an important role in making political discussions more accessible to a broad public.

10.8.2 Rhetoric and Discourse

For good and bad, rhetoric is ubiquitous in democratic discourse. The positive aspect is that it can create interest, understanding and engagement, but there is also a great risk that it is used for biasing messages.

In the first case we can see that good rhetoric follows the theories of Vygotsky 1962; 1978 and Piaget 1962 by adapting a message to the recipient’s zone of proximal development: for pedagogical purposes the message is delivered in a way that will fit the recipient’s conception of the world. This could be regarded as the declarative purpose of rhetoric.

The second case, that of persuasive rhetoric, is arguably more commonplace in politics and serves the purpose of blocking the autonomous thinking of the recipient. This can be interpreted as distrust in the recipient’s ability to make the “right” decisions in complex situations but a more modest suggestion is that it is part and parcel of political discourse.

There are vivid examples of deliberate use of manipulative phrasing. One that has received public attention is Newt Gingrich’s memo from 1996: Language: A Key Mechanism of Control (GOPAC 1990), in which campaign workers are instructed to use certain loaded words about the opposition. However, I believe that it is possible to construct strong arguments without resorting to the use of persuasive language.

10.8.3 Countering Undemocratic Dialogue

Even though I am claiming that democracy is essentially dialogue, not everything that appears to be dialogue is democratic. I will therefore give some examples of what I consider harmful for democratic dialogue.

**Censoring, Ridiculing, or Ignoring**

The greatest threat against democracy is routinized dismissal of dissent. When the discourse becomes biased to regard certain interpretations as true, and op-
position toward them as false, the result is hegemonic dogmatism. This increases the risk of growing silent discontent.

This pattern can be observed in totalitarian regimes, where opposition has been suppressed, but eventually gained enough traction to lead to revolutions. However, many of such revolutions have led to new totalitarian regimes, rather than well-functioning democracies. History suggests that suppression only works temporarily. The threat is that when suppression of dissent bursts, there is a great risk that it will lead to rash, unreflected action and adoption of new dogmas and taboos, and dissent against the new order is yet again suppressed.

Lack of dialogue creates sectarian, polarized groups (Sunstein 2002). This can be countered if dissent is not denied but, instead, challenged. It is not easy, as it presupposes great trust in democratic principles, and also willingness to adjust one’s own position in order to achieve constructive dialogue. Sunstein (2002) makes a strong case for allowing group deliberation in homogeneous groups, as there is a risk that opinions of marginalized members are suppressed in heterogeneous groups. Nevertheless, he also stresses that in order to avoid polarization, such groups need to be exposed to arguments from a wider pool than what is represented within the group.

Misinterpretation
Even though misinterpretation could be addressed in the previous paragraph, it deserves a heading of its own. The reason is that it can occur because of two distinct fallacies or, which is most likely, a mixture of the two. The first is deliberate misrepresentation; when a protagonist wishes to efficiently affect popular opinion by rephrasing the antagonist’s arguments so that they seem more unsympathetic than they were intended (Schopenhauer 2009).

In its purest form, this type of argumentation fallacy is referred to as straw man. The second fallacy is to fail to understand the gist of an argument and thereby attacking the own misconception of it. This is a communication problem that can happen due to a number of reasons, from differing world views and value sets to semantic problems. The latter type of misinterpretation, although sometimes difficult to identify, is not appropriate to classify as undemocratic dialogue but could nevertheless be handled by the same means as the first.

Appeal to Moral Authority
If we accept that humans interpret the world subjectively, we also need to accept that it is impossible to be completely unbiased in the pursuit of moral rights. We, thus, need to accept that every argument—even every process—will reflect the values of its advocate. Reflected or not, some values are sometimes regarded as non-negotiable (Baron and Spranca 1997), which, if confronted with opposing views, leads to fixed positions.

A good example of this type of conflict is different attitudes toward fetal abortion. On one side are those who will not trade off the fetus’ right to be
born, and on the other side are those who will not trade off the woman’s right to decide over her own body. The two moral rights involved are ranked differently by the two factions, and the respective rankings appear self-evident to both sides. There is no way of logically proving that either of them is more right than the other. This kind of disputes cannot be resolved but it is possible to improve the dialogue by creating an understanding of the opponent’s value system, which can lead to acceptance, although not necessarily to agreement. The increasing acceptance of, e.g., same-sex marriage indicates that people are, in fact, prepared to adjust their positions in face of arguments.

10.8.4 A Mind Set for Democratic Dialogue

All of the above, taken together, suggest that dialogue should probably be focused on creating mutual understanding, rather than finding immediate solutions. However, the nice dispositions of someone participating in a democratic dialogue, that are assumed above, can be challenged if we follow the reasoning of Erich Fromm (2001). In *The Fear of Freedom* it is suggested that people will voluntarily refrain from autonomous thinking in face of the uncertainties and unclear responsibilities that follow. Fromm writes that “we have become automatons who live under the illusion of being self-willing individuals” (Fromm 2001, p. 218), by which he wants to point out that the ends that we pursue are not necessarily ours, but more likely transferred to us by socialization. Humans are social creatures and as such we need a social contexts to guide our pursuits, which naturally poses boundaries for what we want to do, and the desires that we have are all reflections of these social contexts. Thus, in face of a pluralistic flood of ideas, we will want to position ourselves where we feel safest, i.e., where we identify our social context to be, and we need to do this in an even firmer way if we perceive a threat by an outgroup.

In order to address this reasonable observation, we cannot simply assume that people unconditionally want to be autonomous. There has to be social support for this disposition. One way to establish this, is by training people to reflect together. For this, computerized tools that support and encourage autonomous reasoning can be valuable. This is the purpose of ColLab.
Part IV:
The End

For the ending of this thesis, I find it appropriate to introduce it in the same way as my PhD studies began. The last words in my Master thesis consisted in a quote from Plato’s *Apology*.

So I left him, saying to myself, as I went away: Well, although I do not suppose that either of us knows anything really beautiful and good, I am better off than he is—for he knows nothing, and thinks that he knows. I neither know nor think that I know. (Plato 2009)
11. Summary

Psychological studies have shown that the past experiences that people value the most are the ones when they felt self-esteem, autonomy, competence, and that they could relate with others (Sheldon et al. 2001). Accepting these findings implies that design should be aimed at increasing, at least not reducing, the possibility to fulfill these values.

Some user experience design theorists have started addressing the question of how to accomplish this (Hassenzahl 2008). This is also my ambition, albeit starting from a moral point of view, namely by focusing on the autonomy of the individual. As Immanuel Kant put it: “But that the principle of autonomy in question is the sole principle of morals can be readily shown by mere analysis of the conceptions of morality” (Kant 1785/2004).

Starting from this assumption, I will argue for why it is important to design for individuals’ autonomy, and why it is important to grant the same autonomy to designers. The definition of autonomy used here differs slightly from the colloquial use of the word, as freedom from coercion, which is essentially a perspective from the outside. In the following, autonomy is instead regarded from the inside: as the ability to make reasoned, rational, purpose-oriented and holistic decisions, with the implication that this is a requirement for being able to take responsibility for these.

11.1 Introduction to Autonomy

Following the argumentation of Kant, I regard autonomy as an unattainable ideal mental disposition\(^1\); as a noumenon. We can approach it but we simply cannot conceive it in its pure form. Our possibility to perceive and manipulate the world is constrained by our physical abilities, as well as by cognitive and dispositional factors.

In the case of cognitive factors, it is useful to think of two different cognitive processes. These are often named System I: the instinctive, automatic,

\(^1\)The question then, ‘How a categorical imperative is possible,’ can be answered to this extent, that we can assign the only hypothesis on which it is possible, namely, the idea of freedom; and we can also discern the necessity of this hypothesis, and this is sufficient for the practical exercise of reason, that is, for the conviction of the validity of this imperative, and hence of the moral law; \textit{but how this hypothesis itself is possible can never be discerned by any human reason.} On the hypothesis, however, that the will of an intelligence is free, its autonomy, as the essential formal condition of its determination, is a necessary consequence.” (Kant 1785/2004, emphasis added)
and effortless reactions to stimuli; and System II: the calculative, logical, and
effortful computational process that we employ when we think hard about
something. If we cannot employ System II thinking, we have no chance to
make reasoned decisions.

In the case of dispositional factors, a dimension emerges in the attitude
toward norms and rules. Jean Piaget wrote: “All morality consists in a system
of rules, and the essence of all morality is to be sought for in the respect
which the individual acquires for these rules” (Piaget 1932, p. 1). On the one
extreme end, we place complete obedience to any command that is imposed on
the individual, whether this derives from external or internalized rules. On the
other end, we have complete independence from such influence. Piaget used
this spectrum to understand the moral development of children: “[...] the main
point is to find out whether one may legitimately alter rules and whether a rule
is fair or just because it conforms to general usage (even newly introduced),
or because it is endowed with an intrinsic and eternal value” (Piaget 1932, p.
15).

Complete obedience cannot be seen as autonomous, but neither can com-
plete independence unconditionally be regarded as that. Wreaking havoc based
on emotional reactions is not deliberate in the way that signifies an autonomous
individual. Furthermore, System II thinking that is fixated by a set of rules
cannot be considered as anything but mere computation. A person that is rea-
soning systematically with the only purpose to follow a prescribed procedure
as well as possible, while completely disregarding the possible consequences,
is not enacting her autonomy and could, without any loss of performance,
be replaced by a machine. Autonomous thinking thus has to be a combina-
tion of System II thinking and a purpose-oriented independence of established
dogmas. It is investigative, reflective, rational, and systematic thinking, as op-
posed to heteronomy, which is unquestioning, habitual, and ritualized. When
you are autonomous, you decide by yourself where to focus attention, and can
motivate your decisions.

In a formulation of his categorical imperative, Kant stressed the responsi-
bility that an autonomously acting agent has: “[Decide] not on account of any
other practical motive or any future advantage, but from the idea of the dignity
of a rational being, obeying no law but that which he himself also gives” (Kant
1785/2004).

Within system development, the agile philosophy (Beck et al. 2001) is cur-
rently very popular and it does, at least in theory, give a high degree of freedom
under responsibility to designers and developers (DDs) (Paetsch, Eberlein, and
Maurer 2003). However, it is not explicit about the value of autonomy, so ef-
fects for users of the produced artifacts cannot be guaranteed. For this latter
aim, one could argue that value-sensitive design methods (B. Friedman and
Kahn Jr. 2003; B. Friedman, Kahn Jr., and Borning 2008) fill the gap, as au-
tonomy is one of the promoted values, but it is unclear how this is accom-
plished. To conclude, it can be argued that although DDs and users sometimes
do enjoy autonomy, it is not prioritized in current practice. Below I will argue for the potential in changing this.

11.2 Allowing Autonomy

If we want people to decide in a more autonomous way, we approach a curious dilemma: any attempt to stimulate autonomous thinking means exerting outside influence, which is inconsistent with the notion of the construct. Furthermore, most everyday decisions are routine matters, in which effortful autonomous thinking would be wasteful, but it is impossible to predict when it is not. Consequently, we should choose the alternative approach, which is to remove conditions that reduce our possibility to use autonomous thinking when that is called for:

If we cannot fully comprehend the situation that we are deciding in so that we need to rely on authorities or simply guess, we cannot be fully autonomous. Neither can we be that if we are not free to decide whatever we rationally conclude to be best.

To design artifacts in an ethical way is important, especially if we consider design as an activity of changing the way people do things. By that current practices are being disrupted. We thus need to take into account the “kingdom of ends”, i.e., people who deserve to be treated as autonomous, in the making of things. We need to assume that they have the ability to be rational, and give them the possibility to be responsible for their decisions.

The implementation of these objectives does not differ much from most of what is considered as good design already today. Usability principles still apply, but are complemented by a goal that is not subjectively given by either designers or users: The complexity of an artifact is determined by how much autonomy the user needs for deciding responsibly. This adds the dimension of appropriateness to the assessment of design complexity. This is particularly important for expert systems. If the reason for using an artifact is to increase the ability to take responsibility, actual complexity cannot be hidden from the user. Instead, she has to be given means to handle it. Such means include both information and stimulation of System II thinking. The main consideration should be to enable users to form proper mental models, both of the problems that they are facing and of the tools that they are using. They must be allowed to gain expertise by using the artifact, so it is crucial to design for controllability and continuous learning. The artifact thus should never prevent users from utilizing their skills nor limit their growth of understanding.

Autonomous thinking is necessary for solving problems that require thinking outside known sets of rules, and autonomy is necessary for being able to take responsibility. Given that I claim this to be ideal for users, it would be inconsistent to constrain the autonomy of DDs by forcing them to follow some rigid method that would ensure user autonomy. Rather, the possibility to be
responsible should be granted also to them. There are reasons for adopting the below described protocol, even if autonomy is not accepted as the highest value. DDs recurrently fail to produce artifacts that satisfy the intended users’ needs, simply because they are addressing the wrong problem; they do not understand what they are expected to produce. The proposed protocol can improve the chance for solving the right problem.

11.3 Theory of Use Protocol

One way to motivate DDs to decide responsibly is to follow a protocol in which they are explicitly held accountable for their designs. However, for this to be fair, they need to be given all means possible to completely understand the problem that they are solving; they need to be allowed to form a proper mental model of the use situation and the users’ needs. They also need to enjoy full freedom to make design decisions founded in rationality rather than in imposed rules. Most importantly, the protocol needs to be forgiving to errors, both because stress inhibits the ability to employ System II reasoning (Kahneman 2012), and because errors are falsifications of current understanding: by exposing misunderstandings, thinking can be corrected.

To accomplish this, I suggest a deductive theorizing protocol based on falsification, inspired by Karl Popper (Popper 2005). It requires a small shift in the role that DDs take in communication with users; rather than considering users as objects of study, the DDs’ understanding is being scrutinized, by users.

When I below write about theory, I mean an explanation for why a model describes what it intends to describe. I use the term in a wide sense, by that disregarding the scientific tradition in which theories emerge only from corroborated hypotheses. By the term model, I mean a description of a system and its mechanics, and by hypothesis, I mean a specific prediction that is motivated by current knowledge, i.e., theory.

11.3.1 Find Meaningful Idea

Popper wrote that “every discovery contains ‘an irrational element’, or ‘a creative intuition’” (Popper 2005, p. 8). For design to be relevant, there first has to be a problem that requires solution. All problems cannot be identified scientifically, especially those that require a level of novelty, as that relies on the ability of the DDs to notice flaws and imagine different futures. In this process, knowledge about the human mind, informed by psychology and philosophy, can guide the focus of DDs. The most important ability is, however, this elusive one called creativity, which presupposes autonomous thinking (Sheldon 1995).
The principle of autonomy that I detail in the thesis can be used to identify a set of problems, through observing the status quo and paying attention to how autonomy is traded off for other values, especially convenience.

11.3.2 Theorize

“From a new idea, put up tentatively, and not yet justified in any way—an anticipation, a hypothesis, a theoretical system, or what you will—conclusions are drawn by means of logical deduction” (Popper 2005, p. 9).

The logical deduction means in this case to work systematically with explicating and exposing the understanding that the DDs have of the use situation; their so-called mental model.

I call this explication a Theory of Use and propose a minimalist structure where the identified problem is first introduced, followed by an account of what DDs consider to be qualities of a good solution, and the assumptions that they are making about the use situation and the users. The latter represents a model of the human-artifact system.

These qualities and assumptions should be formulated as statements that are falsifiable by the users. Presenting these to the users is a systematic way for DDs to get feedback from users on how well they have understood what they are designing for.

The difference between a Theory of Use and a requirement specification is in the focus. Where requirement specifications to a large extent focus on functional requirements and features in a system, the Theory of Use contains a model of the situation in which the artifact is to be introduced and focuses on qualities, in terms of fulfilled needs and goals, in a good solution.

11.3.3 Let the Intended Users React

Falsification means to subject a hypothesis to test in the situation that it concerns. The scientific way to approach this is to reformulate hypotheses so that they predict an effect in the dependent variable if some independent variable is manipulated. In this phase of the theorizing process, the statements in the Theory of Use can be seen as independent variables with the predictions that the intended users will accept them. By this, the DDs put their understanding of the problem up to test by letting users react on it.

The users can, however, not be expected to have reflected on their needs or work procedures, so it is very important that statements are formulated contextually, in a language that the user understands, even if they were derived from high-level abstract theorizing. For instance: “The task requires holistic knowledge of the logistic processes in the system and its current state” can be reformulated as, for instance: “To do your job well, you need to know where the trucks are and what routes they can take”.

165
The Theory of Use should be revised and again subjected to scrutiny by users until it is accepted in its entirety. This phase of the protocol is the most important one, as it develops the DDs’ understanding of what is the problem to be solved and what a good solution satisfies.

11.3.4 Make Hypotheses

After the users are confident enough in the DDs understanding of their problem, as evidenced by the Theory of Use, it is finally time to start creating alternative designs. These should be treated as hypotheses rather than as solutions, which liberates the DDs to use their creative problem solving ability to address the assumptions and qualities in the Theory of Use. As the hypotheses are put up for falsification—by the users for which they are designed—a single one of them is less of a commitment as in current systems development practice, in which DDs are demanded to make decisions on behalf of users. This allows novel, innovative ideas to be evaluated. The general formulation of hypothesis testing in this phase of the protocol is the claim: “This artifact solves your problem”.

When hypotheses are presented to users, it will likely become clear that some assumptions were not accounted for and that others turned out to be wrong, which means that the Theory of Use needs to be revised. Hypotheses work as probes into possible futures, to inform the understanding of both DDs and users. It is also likely that misinterpretations occur in the forming of hypotheses from the Theory of Use. To rectify such errors, it is valuable to be able to point to the agreed-upon assumptions and qualities. As the Theory of Use encompasses the information from the hypotheses, the latter are easy to reject.

11.4 Risks

Even if autonomy is regarded as a necessary value, there are several risks following from increasing it, especially if it is done without solid theoretical foundation. Each of these will turn to actual problems if they are not addressed appropriately. Below, I account for two of these.

11.4.1 Stress

Being explicit about one’s own understanding—thus accountable for it—may result in increased stress. This is particularly true if the motivation is not to increase the chance of designing the right solution but only to expose inabilities of DDs. For the protocol to work, DDs and users need to realize that the groups have respective responsibilities: for the DDs to clearly express
their current understanding, and for the users to assess and rectify this understanding. The responsibility for correctness is thus shared between the groups, which is only effective if the DDs base their later hypotheses on the mutually agreed-upon assumptions and qualities that are stated in the Theory of Use.

11.4.2 Errors

If there are no strict rules to ensure that correct decisions are made, one may fear that a lot of errors may result. It may thus be tempting to reduce autonomy for the sake of correctness. For some types of work, this may be the approach to take, but if the work involves any amount of uncertainty, or demands creativity, it will likely be suboptimal. The proposed protocol is designed not to be robust but to make it easy to recover from errors. If a hypothesis proves to be wrong, it is being resolved either by revising the Theory of Use or by identifying weaknesses in how the theory has been operationalized as hypotheses. Both types of errors yield information about how DDs understand the problem that they are working on.

11.5 Motivation for Thesis

Self-esteem requires the possibility to decide responsibly. Competence can only be achieved through internalization of well-functioning practices, which in turn can only result from a proper understanding of problems. Relatedness requires being respected by other thinking, judging individuals, not just accepted by automatons. All of these values require autonomy, as described above. It is a holistic construct that encompasses that which we value in humans.

Current design practice does not take this explicitly into account. In this summary of my thesis, I have presented a protocol that allows more autonomy for designers and developers. It is intended to improve communication with users, targetting the most important facet of design: the designer’s understanding of what she is designing for.

Moreover, the protocol puts users in control over what is developed for them, and makes them the judges over how to design their lives.
12. Contributions

Most of this thesis is a philosophical discussion. I would argue that this, in itself, is one contribution to research. Much of the philosophy that has been picked up by the HCI community is, for natural reasons, concerning artifact use more than artifact purpose; how more than why. In my opinion, researchers within the field of HCI need to become more explicit about what their motives are, and what moral positions they take when they suggest designs. By offering arguments for one minimalist design philosophy, I wish to spark a discussion of possible alternatives.

Another contribution, which is a natural consequence of the first, is the attempt to synthesize concepts from cognitive psychology with concepts from philosophy.

A third contribution is the protocol that is a translation of the scientific method into practical application. This protocol is also consistent with the design philosophy.

The major contribution is, however, the design philosophy itself. It supplies a means of assessment of design decisions. Questions like whether or not to allow a user to perform X, or what level of abstraction is appropriate, can be addressed on a philosophical basis. By design philosophy, I do not mean only a consistent aesthetic, like simplicity (Colborn 2011), I do not mean only a political ideology, like in Adversarial Design (DiSalvo 2012), and I do not mean only a democratic design process, like in Participatory Design (Ehn 1988). I mean an idea about why artifacts are created and why they should be designed in a certain way, based on an idea what it is to be human. From this follows ideals also for the previously suggested aspects. As both designers and users are currently human, the philosophy should permeate the design process, not only the output of it. Design is inevitably political, as it affects people. A philosophy helps us understand what to expect and supplies arguments for demanding better. Finally, the main issue is not about consistency in design, it is about consistency in thinking.

Although I argue for the universality of the particular philosophy that is offered here, the more important matter is that there should always be an explicit philosophy behind design decisions. This allows a user, or an analyst, to determine a priori if a certain tool is worth considering for a certain application. They can also evaluate in use whether it meets the espoused philosophy. Consequently, analysis can be shifted from evaluating particular features to evaluating overall purposes.
13. Future work

Based on what has been learned in the project about ethical decision making, a new hypothesis for how to motivate people to engage in democratic dialogue is being developed. This work needs to be continued. Other interesting directions for future research based on this thesis include the following:

Apply the Theory of Use protocol in actual development projects. It has taken time to develop a theoretically justified approach to artifact design but, in order to be able to claim with some certainty that it works as hypothesized, it needs to be used in practice.

Evaluate existing artifacts based on the theoretical framework presented in the thesis. A hypothesis to guide this type of evaluation is that if the artifact functions as a tool for the user, it will also be perceived as more useful.

Evaluate the impact of designing for autonomy experimentally. This can be done by creating functionally equivalent artifacts in which specific attributes are manipulated. For instance, the mapping between the user interface and the technical functionality can be distorted to different degrees, in order to evaluate how this affects the understanding of the tool. Another example is to design a decision making tool to evoke emotional reactions, in order to determine how this affects the ability to make unbiased judgments.

Explore how the algorithmic character of digital material can be brought forth in user interfaces (See Chapter 4.3). This is essentially data visualization; not in the sense that the digital is represented perceptibly, but in the sense that algorithms, as abstractions of machine code, become indivisible from their perceptible representation. Apart from that it makes it clear that the digital affords crafting, the clear benefit from this is that it enables diagnosis of the artifact’s actual state.
First of all, I must admit that it has been incredibly difficult to write this thesis. The concepts in it have matured over a long time, and in my thinking it has been obvious how they are related. To turn this tacit understanding into a sequence of words has, however, been the greatest challenge that I have ever been subjected to. To use the terminology in the thesis, I have had a hard time to explicate my mental model.

I could not have done this without the help of many people. Here I will acknowledge a few of those who I have had the honor to discuss ideas with.

First, I am grateful toward my supervisor, Iordanis Kavathatzopoulos, for trusting my ability to accomplish this, and for introducing me to the philosophy and psychology of autonomy. I am a product of your way of thinking; this thesis would not have existed if it was not for you.

My assistant supervisor, Anders Jansson, has read and given valuable feedback on both drafts and the final version of this thesis, which has greatly benefited both my understanding and the presentation of what I am proposing.

My assistant supervisor, Erik Borålv, has given great suggestions for how to get my message across, based in his extensive knowledge about how things are handled in practice. I really appreciate the many other conversations about various topics, including, but not limited to, music, food, and beer.

My colleague Mats Lind deserves a medal for being able to make sense of the first, horrible, draft that I asked him to review. Your critical reading and encouragement has been very helpful and I am really happy that you have joined our group.

I will be eternally grateful to Virginia Grande for proofreading and helping me structure the thesis into a coherent whole. I could not have finished this without your help. I look forward to reading your own thesis!

I am very grateful also to Anton Axelsson for his critical reading and honest comments. Your suggestions for what can be excluded from the thesis were very valuable. I especially appreciated: "this is almost like a bachelor thesis". You are awesome.

The same epithet applies to Jens Meder. I appreciate all your valuable feedback on the thesis, and I am very happy that you see tool value in the presented ideas. I anticipate great things from you.

Apart from Jens, I have had inspiring discussions with many of the HCI master students. I would like mention especially Marten Biehl, Bastiaan Boel, Benjamin Langlotz, and Niknam Moslehi. It has been a pleasure to learn together with you.
Mareike Glöss and Patrick Prax deserve credit for setting very high standard for PhD students. You are both destined to do great things. This applies also to Patrik Stensson, who is finishing a thesis based on the same observations as mine, but from a slightly different angle and with a different target audience. I cannot wait to read it!

I am grateful also to all present and former colleagues at the Department of Information Technology. It is a great place to work, with many warm-hearted and cool people. Special mentions: Rebecka Janols for all the endless discussions about the same few topics, Pontus Olsson and Fredrik Nysjö for knowing everything worth knowing about hardware technology, Fredrik Wahlberg and Cris Luengo for being so kind and insightful, Jonatan Lindén and Nikos Nikoleris for being so handsome, Elisabeth Linnér for being so brave, Anette Löfström, Arne Andersson, Bengt Sandblad, Håkan Selg, Ida Bodin, Lars Oestreicher, Magnus Larsson, Mikael Erlandsson, Ryoko Asai, Thomas Lind, and Åsa Cajander for your excellent choices of research topics and for contributing a lot to my learning. Lena Nordström, Ingela Nyström, Anneli Folkesson, Ulrika Jaresund, Lotta Lundell, and all other heroes who keep the place running. The last colleague to be acknowledged in this, somewhat arbitrarily assembled, list is David “Lök” Eklöv for all the inquiring discussions about politics and economics. These discussions have helped me clarify my thinking and realize in practice how valuable it is to explore conflicting positions if you want to understand your own. We probably still do not agree on many issues, but at least we understand better what the alternatives are.

These acknowledgments would be incomplete if I did not show gratitude toward all my friends, especially toward Eek-a-mouse and Winko. Music made us.

Finally, stort tack to my children Viggo, Edit, and Leif for being patient when their father has not paid enough attention. And Takako, I am amazed by your patience and never-ending support. Thank you for taking care of me and our family. I love you!
Psykologiska studier visar att de erfarenheter som människor i efterhand värdesätter mest är de där de upplevde sig vara autonoma och kompetenta samt när de kände sig bekräftade av andra. Med utgångspunkt från dessa rön ligger det nära till hands att dra slutsaten att vid design av tekniska artefakter bör man sträva efter att öka, eller åtminstone inte minska, möjligheten att uppfylla dessa övergripande värden.


Givet detta antagande kommer jag att argumentera för varför det är viktigt att utforma artefakter för att stödja individers autonomi och varför det är viktigt att samma autonomi omfattar även de som utformar artefakterna.

Definitionen av autonomi som jag bygger mina resonemang på skiljer sig något från den vardagliga användningen av ordet, som mer syftar till frihet från tvång. Det senare är ett utifrånperspektiv medan jag i det följande istället betraktar autonomi från insidan: som förmågan att fatta motiverade, rationella, syftesmedvetna och holistiska beslut. Detta leder till slutsatsen att autonomi är en förutsättning för att i stunden kunna ta ansvar för sina beslut.

14.1 Introduktion till autonomi

Vår möjlighet att uppfatta och manipulera världen begränsas av våra fysiska förmågor men kanske ännu viktigare är att kognitiva faktorer och förhållningssätt begränsar vår förmåga att tänka på ett reflekterande sätt.

När det kommer till de kognitiva faktorerna är det vanligt inom psykologin att modellera två olika kognitiva processer. Dessa kallas ofta för System I—de instinktiva, automatiska, och snabba reaktionerna på stimuli—och System II—den beräknande, logiska, och mödosamma process som vi använder när vi analyserar något. För att att fatta välgrundade beslut måste vi ges möjlighet att aktivera System II.

Vad gäller förhållningssätt framträder en dimension i inställningen till normer och regler. Utvecklingspsykologen Jean Piaget hävdade att all moral består i ett system av regler och att kärnan i all moral står att finna i den respekt som


I en av formuleringarna av sitt kategoriska imperativ betonade Kant det ansvar som en autonom individ har. Man ska inte besluta på grund av praktiska motiv eller framtidiga fördelar utan grunda beslut i idén om värdigheten hos en rationell varelse som inte lyder någon lag som hen inte själv upprätthåller.

14.2 Att tillåta användare att vara autonoma


Om vi t ex inte helt kan förstå beslutssituationen, så att vi måste lita på auktoriteter eller helt enkelt gissa, så kan vi inte vara helt autonoma. Vi kan inte heller vara det om vi inte har frihet och möjlighet att besluta vad vi genom rationellt övervägande kommer fram till är bäst.

Att designa artefakter på ett etiskt sätt är viktigt, särskilt om vi beaktar att syftet med design är att förändra människors sätt att göra saker. Genom detta kan vi åstadkomma omvälvande förändringar i befintliga praktiker. Vi måste därför ta hänsyn till ”kungariket av mål”, som Kant kallar det, det vill säga människor som förtjänar att bemötas som autonoma, när vi skapar och utformar artefakter. Vi måste anta att dessa ”mål i sig själva” har förmågan att vara rationella och ge dem möjligheten att ta ansvar för sina beslut.

Att tillåta autonomi innebär inte nödvändigtvis stora skillnader mot mycket av vad som anses som god design redan i dag. Användbarhetsprinciper gäller fortfarande men kompletteras av ett mål som inte är subjektivt givet av antingen formgivare eller användare: Hur komplex en artefakt bör vara och hur mycket kontroll över den en användare behöver bestäms utifrån vad som ger användaren autonomi. Detta utökar dimensionen för ändamålsenlighet i bedömningen av artefaktens komplexitet och styrbarhet.

Om nivåerna av komplexitet och kontroll är för låga, blir användaren maktlös, och om de är för höga, kommer han ha svårt att utveckla tillräcklig ekspertis. Effekten är dansamma: han kommer inte att kunna ta ansvar för sina beslut.

14.3 Att tillåta designers och utvecklare att vara autonoma


14.4 Protokoll för att teoretisera om användning


För att åstadkomma detta föreslår jag ett deduktivt teoretiserande protokoll baserat på falsifiering, inspirerat av Karl Popper. Det kräver ett litet skifte i rollen som DUer har i kommunikationen med användarna: snarare än att betrakta användare som föremål för studier, är det DUernas förståelse som granskas av användarna.

När jag nedan skriver om teori, menar jag det som en förklaring till varför en modell fungerar som det påstås. Jag använder termen i vid mening, och bortser därför från den tradition där teorier uppstår endast ur hypoteser som bekräftats. Med uttrycket modell menar jag en beskrivning av ett system och dess mekanik och med hypotes menar jag en specifik förutsägelse som motiveras av nuvarande kunskap, dvs. teori.

175
14.4.1 Hitta en meningsfull idé


Principen om autonomi som jag detaljerar i avhandlingen kan även den användas för att identifiera ett antal problem genom att iaktta nuläget och uppmärksamma hur autonomi växlas in mot andra värden, och då speciellt mot bekvämlighet.

14.4.2 Teoretisera

Enligt Popper innebär kritisk prövning av teorier att använda logiska härledningar för att dra slutsatser från en ny, preliminär idé, som kan vara en förväntan, hypotes eller ett teoretiskt system. Den logiska deductionen innebär i detta fall att arbeta systematiskt med att explicitgöra och exponera förståelsen som DUerna har av användningen och situationen, deras så kallade mentala modell. Utifrån förståelsen formuleras falsifierbara påståenden.

Jag kallar detta explicitgörande för en teori om användning och föreslår en minimalistisk struktur där det identifierade problemet först introduceras, följt av en redogörelse för vad DUerna anser vara egenskaper hos en bra lösning samt de antaganden som görs om användningssituationen och användarna. Det senare är en modell av människa-artefaktsystemet.

Dessa egenskaper och antaganden bör formuleras som uttalanden som är möjliga för användarna att falsifiera. Språket bör vara sådant att det är meningsfullt för både DUer och användare. Att presentera antagandena för användare är ett systematiskt sätt för DUerna att få återkoppling på hur väl de har förstått vad de designar för.

Skillnaden mellan en teori om användning och en kravspecifikation ligger framför allt i fokuset. Där kravspecifikation till stor del fokuserar på funktionella krav och funktioner i ett system, innehåller teorin om användning en modell av den situation i vilken artefakten ska introduceras och fokuserar på kvaliteter i en bra lösning, med avseende på hur användarnas behov och syften uppfylls.

14.4.3 Låt tänkta användare reagera

Falsifiering innebär att utsätta en hypotes för prövning i den situationen som den avser. Det vetenskapliga sättet att närma sig detta är att omformulera
hypoteser så att de förutspår en effekt på den beroende variabeln om någon oberoende variabel manipuleras. I denna fas av teoretiseringsprocessen kan uttalandena i teorin om användning betraktas som oberoende variabler med hypotesen att de tänkta användarna kommer att acceptera dem. Genom detta sätts DUernas förståelse av problemet på prov genom att låta användarna reagera på den.

Användarna kan dock inte förväntas ha reflekterat över sina behov och arbetsrutiner så det är mycket viktigt att påståenden formuleras kontextuellt, på ett språk som användaren förstår, även om de är sprungna ur ett teoretiskt resonemang. Till exempel: "Arbetsuppgiften kräver holistisk kunskap om de logistiska processerna i systemet och systemets status" bör formuleras till exempelvis: "För att göra ditt jobb bra måste du veta var bilarna är och vilka vägar de kan ta".

Teorin om användning bör ses över och återigen göras till föremål för granskning av användare tills det accepteras i sin helhet. Denna fas i protokollet är den viktigaste eftersom det utvecklar DUernas förståelse för vilka problem som ska lösas och vilka värden och syften en bra lösning uppfyller.

**14.4.4 Utforma hypoteser**


När hypoteser presenteras för användarna kommer det sannolikt att bli tydligt att vissa antaganden inte redivisas och att andra visade sig vara felaktiga, vilket innebär att teorin om användning behöver revideras. Hypoteserna hjälper därigenom DUerna att sondera möjliga framtider, och bidrar till både DUernas och användarnas förståelse av problemet. Det är också troligt att feltolkningar uppstår i utformandet av hypoteser från teorin om användning. För att korrigerar dessa fel är det värdefullt att kunna peka på de överenskommna antagandena och kvaliteterna. Eftersom teorin om användning utvecklas genom
den information som kommer fram i hypotesprövningarna, innebär det ingen stor förlust att förkasta hypoteser.

14.5 Risker
Även om autonomi betraktas som ett nödvändigt värde finns det flera risker som följer av att utöka denna, särskilt om det sker utan solid teoretisk grund. Var och en av dessa risker kommer att växa till kritiska problem om de inte behandlas på rätt sätt. Nedan diskuterar jag två av dessa.

14.5.1 Stress

14.5.2 Fel
Om det inte finns några strikta regler för att säkerställa att rätt beslut fattas kan man befara att en hel del fel kan uppstå. Det kan därför vara frestande att reducera autonomi med motiveringen att det inte får gå fel. För vissa typer av arbete kan detta vara rätt men om arbetet innebär någon form av osäkerhet eller kräver någon nivå av kreativitet kommer det sannolikt att leda till suboptimalt resultat. Det föreslagna protokollet är inte konstruerat för att vara felsäkert utan för att göra det lätt att korrigera fel. Om en hypotes visar sig vara felaktig löses det antingen genom att revidera teorin om användning eller genom att identifiera brister i hur teorin har operationaliserats som hypoteser. Båda typerna av fel ger information om hur DUerna förstår problemet som de arbetar för att lösa.

14.6 Motivering till avhandlingsämnet
För att sammanfatta sammanfattningen: Kompetens kan endast uppnås genom internalisering av väl fungerande angreppssätt, vilket i sin tur bara kan följa av en korrekt förståelse av problemen man angriper. För att kunna relatera till andra behöver man uppleva att man respekteras av andra tänkande individer, inte


Baron, Jonathan and Mark Spranca (1997). “Protected Values”. In: *Organizational Behavior and Human Decision Processes* 70.1, pp. 1–16.


Fischer, Gerhard (2002). “Beyond “Couch Potatoes”: From Consumers to Designers and Active Contributors”. In: *First Monday* 7.12.


Greenberg, Saul and Bill Buxton (2008). “Usability evaluation considered harmful (some of the time)”. In: *Proceedings of the SIGCHI Conference*
Gulliksen, Jan et al. (2003). “Key principles for user-centred systems design”. In: *Behaviour & Information Technology* 22.6, pp. 397–409.


Le, Quoc et al. (2012). “Building high-level features using large scale unsupervised learning”. In: International Conference in Machine Learning.


Rauschmayer, Felix et al. (2009). “Why good practice of OR is not enough : Ethical challenges for the OR practitioner”. In: *Omega* 37.6, pp. 1089–1099.
Sussman, Gerald (2011). We Really Don’t Know How to Compute. URL: http://www.infoq.com/presentations/We-Really-Dont-Know-How-To-Compute.
12. Lars Thofelt: Studies on leaf temperature recorded by direct measurement and by thermography. 1975.
19. Thomas C. Amu: Diffusion in Dilute Solutions: An Experimental Study with Special Reference to the Effect of Size and Shape of Solute and Solvent Molecules. 1978.