



UPPSALA
UNIVERSITET

IT Licentiate theses
2010-005

Let Us Be Philosophers!

Computerized Support for Ethical Decision Making

MIKAEL LAAKSOHARJU

UPPSALA UNIVERSITY
Department of Information Technology



Let Us Be Philosophers!
Computerized Support for Ethical Decision Making

Mikael Laaksoharju
Mikael.Laaksoharju@it.uu.se

September 2010

Division of Human-Computer Interaction
Department of Information Technology
Uppsala University
Box 337
SE-751 05 Uppsala
Sweden

<http://www.it.uu.se/>

Dissertation for the degree of Licentiate of Philosophy in Human-Computer Interaction

© Mikael Laaksoharju 2010
ISSN 1404-5117

Printed by the Department of Information Technology, Uppsala University, Sweden

Abstract

This thesis presents a computerized tool for ethical decision making. For someone who is unfamiliar with the psychological theory that the tool is based on, it will perhaps first appear as a pointless piece of software. It does not give any guidance to what an ethically correct decision is, it does not suggest relevant ethical principles or guidelines and it does not even make reference to known cases of good moral conduct. In fact, it does not make any moral claims at all. The only two things that the tool does are that it stimulates reflective, analytical and holistic reasoning and blocks automatic, biased and constrained impulses. This approach is chosen to improve the decision maker's ability to consider the relevant circumstances in a situation. By focusing on relevant interests of stakeholders, the scope of consideration in a moral situation can be expanded and the impact of decisions can be evaluated with respect to these. To justify this non-normative approach, the functionality of normative ethics is analyzed. The conclusion stresses the importance of self-conscious deliberation. Further arguments for advocating a systematic, holistic and self-critical handling of moral problems are collected from both philosophy and psychology. The structure and functionality of the tool is founded in psychological theory and especially the problem of cognitive biases in moral decision making is addressed. The tool has been evaluated in two studies, which both indicate that it actually delivers what it was designed to do. Statistically significant results show that the tool helped users to expand the scope of consideration in a moral problem situation compared to using an equivalent paper-and-pen-based method.

Svensk sammanfattning

Det här är en licentiatavhandling om hur man kan hantera moraliska problem med hjälp av datorstöd. Arbetet bygger på två antaganden. Det första är att moraliska problem uppstår framför allt som en följd av bristande kunskap om betingelserna för en beslutssituation. Det andra är att moraliska problem kan hanteras genom att på ett systematiskt sätt ta reda på hur intressen hos berörda individer och organisationer påverkas av och påverkar ett beslut. På detta sätt kan man skapa sig en uppfattning om de effekter som olika beslut medför. Syftet med denna skärskådning är att blockera ett begränsat, dogmatiskt och automatiserat tänkande för att istället stimulera ett sakligt, reflekterande och opartiskt resonemang. Istället för att förblindas av abstrakta principer och ogrundade föreställningar bör man analysera hur människors konkreta värden och intressen påverkar och påverkas i den aktuella situationen. Det låter självklart och enkelt men är väldigt svårt att uppnå i moraliska frågor som oftare än inte är laddade med starka känslor.

De flesta organisationer har regler och riktlinjer för hur moraliska frågor bör hanteras. Antalet och utformningen av dessa kan variera alltifrån väldigt många, detaljerade regler till några få, övergripande riktlinjer. Sådana regelsamlingar kan vara värdefulla enbart för att de signalerar företagets etiska ambitioner till omvärlden men de bör också hjälpa medlemmar av organisationen att fatta beslut i svåra situationer. Det verkliga värdet av dem kan bara avgöras i konkreta situationer, men problemet med verkligheten är att den allt som oftast inte passar in i riktlinjerna. För att komplettera dessa, när de inte är tillämpbara, behövs därför sunda processer för att hantera moraliska ställningstaganden¹. För detta syfte har jag utvecklat ett datorstöd. Det heter EthXpert och är ett verktyg avsett att hjälpa beslutsfattare med att skapa sig en så korrekt och fullständig uppfattning som möjligt om ett problem med moraliska aspekter². Verktöget stödjer beslutsfattaren genom att visualisera intressena hos de aktörer som påverkar och påverkas av ett beslut. Det riktar sig till en bred publik och förutsätter inte något särskilt innehåll i problemen som ska analyseras. Det refererar varken till tidigare kända moraliska problem eller till normativa principer och pådyvlar således inga pekpinnar på beslutsfattaren. Vad som kännetecknar ett moraliskt beslut kan inte antas vara känt på förhand och därför måste verktöget vara

¹Se paper IV.

²Se paper I.

konstruerat så att den inte gör några bedömningar av den normativa riktigheten i innehållet. Denna avsaknad av riktlinjer kan inledningsvis verka som en svag strategi – hur ska man då veta vad som är moraliskt att göra? Dock visar sig detta vara en styrka när det gäller att utvidga förståelsen för problemsituationen. Användaren lockas aldrig in i en falsk föreställning att analysen är klar bara för att verktyget säger så utan måste själv avgöra, baserat på den egna moralen, när analysen är tillräcklig. Etiska överväganden bör alltid lämna en känsla av osäkerhet och öppna för självkritik, för hur skulle vi någonsin kunna vara säkra på att vi gör rätt när inte ens de stora filosoferna har kommit överens om det? Det enda vi kan göra är att försöka så gott som möjligt och använda den metod som de genom århundraden har tillämpat på moraliska problem, nämligen att systematiskt, kritiskt och opartiskt analysera alla aspekter av en valsituation. En sådan inställning delegerar ansvaret för en tillfredsställande analys helt till den som utför analysen och ökar därmed också incitamenten för att göra rätt.

Verktyget har utvärderats i två studier som presenteras i avhandlingen³. Resultaten är positiva då samtliga deltagare lyckades utöka omfattningen av vilka som påverkar och påverkas av problemen. Det är halvvägs mot ett beslut att inse hur det ingår i en komplex verklighet. Dock återstår en hel del arbete för att göra verktyget bättre på att presentera resultatet av analysen så att beslutsfattare kan tillgodogöra sig hela den komplexitet som utgör en moralisk beslutsituation. Detta är en spännande utmaning som vidare forskning inom ämnet människa-datorinteraktion bör kunna bidra till.

³Se papper II och papper III.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals. Reprints were made with kind permission from the publishers.

| | |
|-----------------|--|
| Paper I | Computerized Support for Ethical Analysis |
| Authors | Laaksoharju, Mikael and Kavathatzopoulos, Iordanis |
| Publication | Maria Botti (Ed.), Proceedings of CEPE 2009 – Eighth International Computer Ethics Conference. Corfu, Greece: Ionian University, 2009 |
| Short summary | The paper describes the functioning of a computerized tool that supports decision makers in their acquisition of relevant values and interests in a problem situation. |
| My contribution | I wrote most of the paper. The paper describes a tool designed and constructed by me. |
| Paper II | Ethical Usability of IT Systems: How to Consider Relevant Factors and How to Find Solutions |
| Authors | Kavathatzopoulos, Iordanis and Laaksoharju, Mikael |
| Publication | Proceedings of the Fifth Asia-Pacific Computing and Philosophy Conference. Tokyo: Tokyo University, 2009 |
| Short summary | The paper gives a philosophical background to the autonomy approach toward moral problem solving and presents an exploratory study where a computerized tool to address moral problems was applied on ICT systems. |
| My contribution | I conducted the study reported in the paper and wrote the parts that describe it and the part that covers how the system functions. |

| | |
|------------------|---|
| Paper III | |
| Authors | Kavathatzopoulos, Iordanis and Laaksoharju, Mikael |
| Publication | M. Arias-Oliva, T. W. Bynum, S. Rogerson and T. Torres-Coronas (Eds.), Proceedings of the Eleventh International Conference: The “Backwards, Forwards and Sideways” Changes of ICT, ETHICOMP 2010. Tarragona, Spain: Universitat Rovira i Virgili, 2010 |
| Short summary | The paper describes a study where a tool for ethical decision making was applied on various moral problems. |
| My contribution | I wrote most of the paper, did the literature review and summarized the qualitative data. The studies, as well as the quantitative analyses, were performed together with Iordanis. |
| Paper IV | |
| Authors | Kostrzewa, Agata, Laaksoharju, Mikael and Kavathatzopoulos, Iordanis |
| Publication | G. J. M. da Costa (Ed.), Ethical Issues and Social Dilemmas in Knowledge Management: Organizational Innovation. Hershey, USA: IGI Global, 2010 |
| Short summary | This book chapter introduces the autonomy approach in Knowledge Management by discussing how it can contribute to utilizing moral knowledge within organizations. |
| My contribution | I wrote the paragraph on EthXpert. The introduction, future research directions and conclusions were written together with Agata. I also contributed to the rest of the paper by careful reading and feedback. |

My Co-authors

Iordanis Kavathatzopoulos, PhD in Psychology. Professor in Human Computer Interaction at the Department of Information Technology, Uppsala University.

Agata Kostrzewa, PhD student at the Department of Business Studies, Uppsala University.

Other publications

Laaksoharju, M. and Kavathatzopoulos, I. (2008). Tools for ethical decision making. In F. Wenstøp and J. P. Brans (Eds.), *OR & Ethics: Human Centered Processes Third International Conference, Part III*. Delft, Netherlands: TELECOM Bretagne (pp. 13-24).

Laaksoharju, M. and Kavathatzopoulos, I. (2008). Can micro world simulations assess and stimulate ethical competence? In T. W. Bynum, M. Calzarossa, I. De Lotto and S. Rogerson (Eds.), *Living, Working and Learning Beyond Technology, 2008: Proceedings of the 10th International Conference*. Mantova, Italy: Tipografia Commerciale (pp. 503-510).

Kavathatzopoulos, I., Laaksoharju, M. and Rick, C. (2007). Simulation and Support in Ethical Decision Making. In T. W. Bynum, K. Murata and S. Rogerson (Eds.), *Globalisation: Bridging the Global Nature of Information and Communication Technology and the Local Nature of Human Beings, ETHICOMP 2007: Proceedings of the 9th International Conference*. Meiji University, Tokyo (pp. 278-287).

Contents

| | |
|--|----|
| Introduction..... | 13 |
| Outline of the Thesis..... | 15 |
| The Research Area of HCI..... | 16 |
| On Computer Ethics..... | 19 |
| Research Project..... | 20 |
| ETHCOMP | 20 |
| ETHCOMP Tools..... | 20 |
| The ETHIX Questionnaire..... | 21 |
| The OLE Questionnaire..... | 21 |
| The P&P Method..... | 22 |
| Ethick Simulator..... | 22 |
| The EthXpert Software..... | 23 |
| Theory..... | 24 |
| On Philosophizing..... | 24 |
| On Moralizing..... | 28 |
| On Morality..... | 29 |
| Cognitive Dissonance | 31 |
| Decision Making Strategies | 32 |
| Heteronomy..... | 33 |
| Autonomy..... | 33 |
| Ethical competence..... | 35 |
| Procedural Ethics..... | 36 |
| On Calculating..... | 37 |
| Related work..... | 39 |
| EthXpert..... | 41 |
| Bias Prevention..... | 44 |
| States of Ethical Analysis..... | 45 |
| 1. Define stakeholder | 47 |
| 2. Define interest for stakeholder | 48 |
| 3. Define how interest relates to stakeholders | 49 |
| 4. Define main option/scenario | 50 |
| 5. Translate considerations..... | 50 |

| | |
|--|----|
| 6. Define compromise option/scenario | 50 |
| Technical description | 51 |
| Methodology and Methods | 53 |
| Usability of EthXpert | 53 |
| Study 1 | 54 |
| Study 2 | 54 |
| Analysis | 55 |
| Results | 56 |
| Satisfaction Ratings | 56 |
| Qualitative Analysis | 57 |
| Positive | 58 |
| Mixed | 60 |
| Negative | 60 |
| Discussion | 63 |
| On Interests | 64 |
| On Stakeholder Participation | 65 |
| Options | 66 |
| Is this moral? | 66 |
| Future work | 67 |
| Acknowledgements | 68 |
| Bibliography | 69 |

Introduction

Let me start with an example. A medical doctor wants to analyze blood from a patient in order to rule out possible causes for the symptoms that she has observed. She therefore sends blood samples to the laboratory along with a list of tests that she wants to have run. In the laboratory, the technician puts the samples into a machine that automatically runs a standard set of tests, far more numerous than the ones requested by our doctor. The policy at the hospital is however to only return the values that were requested by the physician, even if other values would show abnormalities.

It would be easy to condemn this practice as unethical, if there was not the complicating factor that the tests give false positives in 5% of the cases. If all the results from the standard test were returned it would mean that a lot of patients would have to undergo unnecessary further investigations. The expense for the hospital and the discomfort for the patients who are not ill have been determined to be enough reason to not second-guess the doctor's initial judgment, even though it in some cases probably could save lives.

Most organizations have policies and guidelines concerning decisions where moral judgment is required. The above example is made up, but is illustrative for a situation when policies can relieve individuals from the burden of judging in moral issues. In best case these are the result of ethical contemplation, which would mean that ethically competent people have deliberated over moral issues in relevant topics and created codes and principles for how to act morally. This is the ideal, but it is still limited in many ways. For obvious reasons, the nature of the resulting guidelines usually becomes general instead of specific. For a person facing a moral problem, this means that the original problem of deciding what to do in a difficult situation is replaced with the problem of determining which guidelines apply in the current situation. Not to mention the problem of determining *how* they apply.

Moral problems are generally characterized by conflicting values or principles and therefore very hard to deal with, so there indeed is a strong need for guidance. But when deploying a rigid set of guidelines, the focus of problem solving shifts, something that has several undesirable implications: solving moral problems turns into a rule based activity and is thus limited to previously predicted scenarios; the awareness of unique features in specific problems decreases and relevant questions that should be raised about a

problem may end up disguised by untouchable principles and therefore neglected (see for example Eriksson et al., 2007). The feeling of responsibility for a decision can also be reduced when the actual decision has been made in advance by someone else, and then communicated as a list of items to consider. And what happens when the guidelines are not directly applicable to the imminent moral problem? This could leave the individual to base the decision on her own judgment – a judgment that might be constrained by reluctance to act against norms but also biased by personal inclinations and values. Sadly, it is probably more common that the individual is not even aware of there being moral considerations in a case, as the skill of ethical awareness has not been properly trained. The most destructive scenario is that an individual might avoid to see that there exists a moral problem because of the uncertainty and call for responsibility that it causes. Most of us are simply not well prepared to handle moral problems. Whenever we experience that we have two or more conflicting moral obligations to fulfill we become insecure about both our ability and our authority to handle the problem. Moral problems often seem impossible to solve without violating at least one important value. To complicate matters more, some people hold values that they will not trade off for any cost. These are called protected values (Baron and Spranca, 1997). To these people a situation can appear as if there is no dilemma, which can make the problem even harder for someone who has to find a middle way between conflicting protected values. Fetal abortion is an example of this kind of conflict. On one side are the people who will not trade off the baby's right to live and on the other side are the ones who will not trade off the woman's right to decide over her own body. In difficult situations we generally want to think that we are out of options so that we do not have to take responsibility for our decisions. Psychologically we are constituted to avoid situations where we risk losing something, so we try hard in different ways to avoid creating or ending up in such situations. This can in some cases also lead us to the false conclusion that inaction is better than action (Sunstein, 2005). A final complicating factor is that moral problems often regard issues that are somehow sensitive. Consequently there is a risk that our decision making becomes biased by emotions and taboos. In fact, neurological research implies that most of us always involve emotions in moral decisions (Greene and Haidt, 2002; Koenigs et al., 2007). Generally this is something good, as it usually leads to decisions congruent with the society's morality, but it underlines that we are biased and will often decide automatically, without thinking.

The above paragraphs are intended to highlight that, although policies and guidelines are helpful in many situations, other situations cannot be satisfactorily solved without moral judgment. The latter type of situations can be further divided into two types: The situations that can be solved by relying on our immediate moral sensibility and the situations that lead to

terrible consequences if we act according to our gut feeling. The first type are luckily very much more common – morality is indeed a purposeful construction. The only problem is that our moral heuristics do not always ring the bell when we have encountered a situation of the second type.

Outline of the Thesis

When I set out to write this licentiate thesis I had great ambitions. I wanted to tell stories that make people wiser. I did not want the ideas that I was to present to be considered as academic obscurities that would never have any practical relevance. Therefore I started to write short stories that would secretly sneak into people's consciousnesses so that once there, they could explode like Trojan horses and pollute the minds with questions and ideas. Some of these stories were absurd, some were pointless, yet others could have been valuable in some sense, to someone. Further into the process I realized that I could not fulfill the initial ambition. The reason why I sought that approach is the very reason why I could not accomplish it: Morality is difficult to approach analytically since moral topics persistently demand focus on content instead of the way that they are handled.

To easier maintain an analytical perspective, this thesis is therefore a retreat to the standard structure of an academic text. First I will introduce you to the problem and the research conditions, after which I will establish some useful theories to justify the design of the ethical decision support system. The latter half of the thesis describes how these theories have been translated into a computerized tool and how this tool has been evaluated.

I will below argue that a systematic scrutinizing of the interests of stakeholders in situations with ethical considerations can help decision makers to reach ethically better decisions. Furthermore, as the main point of this dissertation, I will claim that this scrutinizing can be successfully assisted by purposefully designed software based on a clear and practical conception of ethics.

To support these claims I will have to suggest an interpretation of ethical decisions that does not rely on universal ethical theories and eternally determined rights and wrongs. I should also show, or at least give strong arguments for, that this perspective will actually be useful in practice.

For the reader whose attention is helped by knowing the objective for a text I will hereby sketch a research question that captures the essence of my work, while leaving enough ambiguities to motivate further reading:

“How should a computerized tool be designed so that it can stimulate decision makers to make morally good decisions?”

The Research Area of HCI

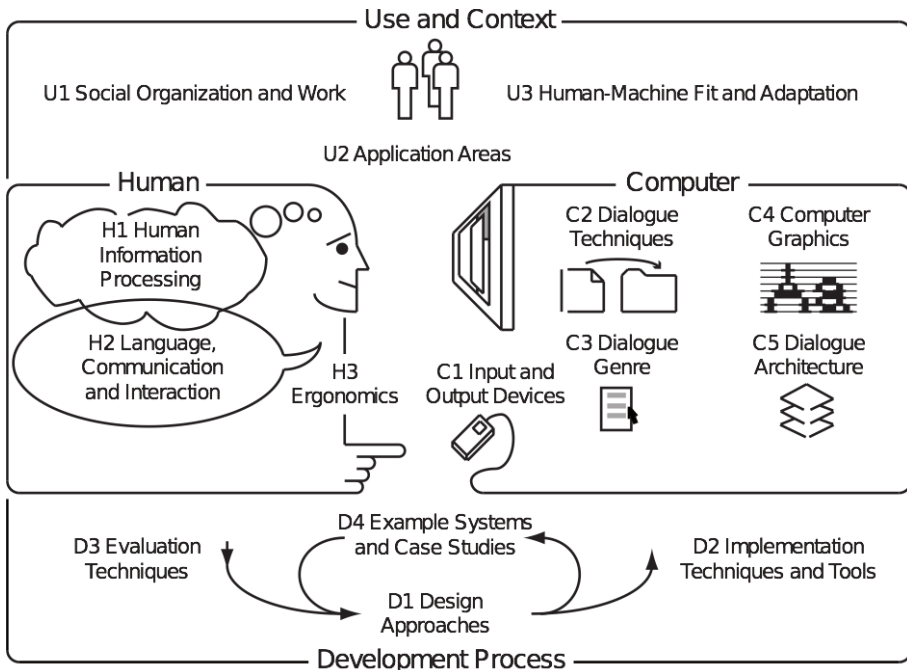


Illustration 1: Typical topics in HCI as described in ACM SIGCHI Curricula for Human-Computer Interaction. Source: ACM, 1992.

“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, 1992)

Illustration 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 Curricula⁴. Indeed it visualizes well the scope of the field but without the supporting manifest it leaves to the viewer or presenter to imagine how the various topics are addressed and,

⁴ACM (Association for Computing Machinery) SIGCHI (Special Interest Group on Computer Human Interaction) Curricula for HCI.

perhaps more importantly, not addressed within HCI. As I curiously went to the source (ACM, 1992) to become enlightened about these constraints I found that practically any research involving humans and computers match the description. In fact, what is today regarded as neighboring research areas, like e.g. 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. Still seven years ago, in the foreword to the Human-Computer Interaction Handbook (Jacko and Sears, 2003), Ben Schneiderman 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 is 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 can mature into a full-fledged basic discipline such as physics or psychology, it can continue to be a inter- and multidisciplinary meeting place and it can 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. In fact, I have become increasingly comfortable with the idea of HCI as purely an interdisciplinary research topic and not at all as a discipline. On the other hand, if it 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 unique enough anymore.

One idea comes from Gilbert 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 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. It may sound like a naive and vague dream but after some reflection this miss-universe-like 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 not enough for HCI researchers to satisfy with ethnographic studies to observe the current situation or 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 current definition!) 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 as an important purpose as facilitator 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 research focus within HCI has changed from time to time – from computer systems thinking over cognitive psychology to ethnography and contextualization (and now we find ourselves in something that in retrospect probably will qualify as the decade of hedonism and content free 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 the establishment of its research, but in order to gain legitimacy it might be necessary to be concise about what HCI research focuses on. 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 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. To sum this up, I will just conclude that HCI researchers still have some disciplinary issues to bring up for discussion. The challenge in the above quote from the ACM SIGCHI Curricula is to define an appropriate scope for “the study of major phenomena surrounding [interactive computing systems]”.

On Computer Ethics

It is not very controversial to claim that ICT comes along with moral problems. There is not much debate over whether there are moral implications in ICT issues or not. Instead, philosophers and other academics are involved in a discussion about how to approach these implications. Some claim that a new research discipline has evolved, covering unique problems that require new theories and methods (e.g. Floridi, 1999; Maner, 2006), while others claim that moral problems within ICT can be treated as any other moral problems. I claim that the evolution of ICT, more than anything else, proves that recipes for how to act morally, e.g. codes of conduct or guidelines, are never sufficient. ICT indeed has created conditions for moral problems that are somewhat unprecedented: Information can be multiplied and spread at practically no cost and can be processed and synthesized in ways that pose threats to our privacy as well as other values, our dependence on infrastructure makes us vulnerable to breakdowns etc. (see e.g. Maner, 2006). However, even though the conditions have changed, we have not – we still act very much like human beings in a human context. More important, though, is that society continuously changes to embrace new technology and therefore also will incorporate its moral problems. I do not claim that Computer Ethics is unnecessary as a research subject – it is of course important that at least some philosophers have domain expertise in ICT. Nevertheless, I do suggest that Computer Ethics (and any other related philosophy), as computers become ubiquitous, will inevitably become incorporated in the scope of “common” macroethics as once did the use of printing press, trains, phones and cars. It is likely, as Floridi (1999) advocates, that ICT and technology will change macroethics⁵, but then again, as will become apparent below, I am not convinced about a wide-spread normative impact of *any* ethical theory and therefore prefer to regard these as descriptive rather than prescriptive.

⁵Floridi is actually proposing to fundamentally change the concept of ethics from a concern for humans (anthropocentric) to a concern for generalized objects (ontocentric). This would make us morally obliged to care for the information found in the infosphere and would also give autonomous robots etc. rights and duties. As we live our lives to an increasing extent online, and thus in a world of information, we will have to develop morality to handle it. An information ethics might be better fit to describe this emerging phenomenon than traditional ethics.

Research Project

ETHCOMP

The project ETHCOMP – Ethical Competence for Decision Makers and Organizations – started in fall of 2007 and will end in autumn 2010. It was funded by Handelsbanken Research Fundings: Jan Wallanders & Tom Hedelius Stiftelse and Tore Browaldhs Stiftelse. The work was based on a specific definition of ethical competence: for individuals as well functioning mental processes in problem solving and decision making, and for organizations as purposeful group processes independent of normative aspects. The project resulted in several tools and methods to promote and support ethical competence.

- Questionnaires for testing, mapping and indexing organizational ethical competence
- Methods to work with ethical guidelines and values
- Training methods to increase the ethical competence of decision makers, e.g. a micro world simulator for ethical problem solving. Used for both assessment and stimulation of ethical competence
- Tools to support ethical decision making, like the tool presented in this thesis

ETHCOMP Tools

The tools and methods developed in the research project, ETHIX, OLE, P&P, Ethick and EthXpert, all support autonomy, i.e. the adoption of a holistic approach, of critical reasoning and of systematic analysis of ethical aspects in decision making. The aim is to help to identify the significant problems, to make them explicit, and to reformulate them in order to be able to work with them. Decision makers define their own (organizational as well as personal) positions, duties, commitments, values and feelings. They also identify and take into consideration the interests, values and needs of stakeholders. They generate alternative courses of action, and they

systematically weigh each of them against all values and interests involved in the situation. By that they can consider important relevant ethical aspects and adopt the ethically most suitable solutions.

Fundamentally, all of the tools are based on the same assumptions. They all advocate focusing on interests, values and needs of involved stakeholders. OLE does this through specifically directed questions, and is, like P&P, convenient, as the only necessary properties are paper and pen or a word processor. EthXpert however has benefits in visualizing the relationships between stakeholders and adding the possibility to easily reuse and restructure information. It also features the additional step of analyzing how the interests and values of stakeholders affect other stakeholders.

The ETHIX Questionnaire

The Ethical Index (ETHIX) is a survey constructed to describe ethical competence in organizations, on both individual and organizational levels (see paper IV). It is aimed at identifying organizational ethical strengths and weaknesses, to map needs and to plan, follow up and evaluate organizational changes. It can be used to assess the effect of organizational actions such as personnel training, introduction of ethical processes and roles, etc. ETHIX summarizes how an organization manages ethical processes on both group and personal levels.

The questionnaire in its full version consists of 37 questions of Likert-type with six alternatives (see appendix, paper IV). The survey has already been tested on two organizations- a university department and a manufacturing company. Parameters of ethical competence at personal level include ethical awareness, problem-solving and decision-making skills, argumentation skills, ethical training, and confidence in making moral decisions. At the organizational level, organizational profile and concern in moral issues, adoption of special processes and roles, and support to the members are important. Another important aspect concerns ethical guidelines: how these are constructed, revised and what impact they have in decision making.

The OLE Questionnaire

The questionnaire on Organization, Learning and Ethics (OLE) poses a set of questions aimed at supporting the identification of important non-technical aspects and problems in IT use. The questionnaire directs attention to specific topics that are likely to cause ethical problems. The assumption is that working with these questions will, in a straightforward way, raise the awareness of the important aspects and facilitate measures for improving usability and productivity.

The P&P Method

The paper and pen method (P&P) integrates two aspects of decision making. The first part is done as a preparation for the second part in order for the participant to be conscious about the difference between different ways of moral problem solving and decision making. It consists of two columns where the participant can give examples of heteronomous (knee-jerk, emotional, dogmatic, uncritical) and autonomous (systematic, critical and holistic) thinking.

The second part is a matrix with values on one axis and alternative courses of actions on the other (Erlandsson and Kavathatzopoulos, 2005; Rauschmayer et al., 2009). The task is to fill in the values relevant for the situation as well as the possible options. The decision maker compares each value with each option, which results in small descriptions for each cell of the matrix. The small summaries describe how each solution affects each value. While doing this, even more values and options will evolve out of the considerations for each cell.

The process of constructing such a matrix can never be completely finished, but the more people involved in the creation the more complete the matrix will become. The resulting matrix does not state any correct solution, but rather what effects each solution can have on all involved values. Therefore, to use this matrix as a decision support one has to make a conscious choice between the identified decisions, without being able to ignore all the positive and negative implications that are clearly stated for each decision.

Ethick Simulator

Ethick (Laaksoharju, 2008; Laaksoharju and Kavathatzopoulos, 2008) is a micro world simulator that I constructed for my master thesis. The purpose of the simulator is to train the user's ability to think autonomously and also allow assessment of this. A micro world simulator is a computerized model of a world in which autonomous (in computer terminology, i.e. heuristic) agents interact in some way. In Ethick, the researcher defines virtual stakeholders and assigns interests and inclinations to them, e.g. hunger, thirst for fame, safety etc. These constitute the primary mechanisms that drive the stakeholders' decision making in the micro world and are subjective to each agent. Outer guides in the form of simple principles like equal distribution of wealth, equal access to certain assets etc. can also be implemented. Within certain boundaries, the stakeholders will then try to maximize their satisfaction in the world. They will consume assets that fulfill their current desires, e.g. food if they are hungry and lectures if they are thirsty for knowledge. If any of the stakeholders obey a principle it will include the welfare of other stakeholders in its decision making.

When a test subject is put in front of the running simulator, her task is to consider the needs of all stakeholders. To affect their behaviors she is encouraged to create rules in the micro world whenever identifying possible causes of conflict. To accomplish this, she is supplied a rule creation wizard that asks questions about various aspects of the intended rule, e.g. which stakeholders or assets or interests to consider and how. Every interaction with the system is being recorded in order to subsequently trace clues about the thought process behind each of these created rules. The hypothesis is that the interaction pattern will covary with achievement in ethical competence tests.

The EthXpert Software

EthXpert is a tool intended to help decision makers in the process of establishing a conception, as complete as possible, about a problem with ethical implications. It is designed to block biased reasoning and support a systematic, self-critical and holistic approach toward moral problems. The procedure focuses on the interests, values and principles of stakeholders involved in the problem situation. This tool will be presented further below.

Theory

On Philosophizing

Societies constitute systems. These systems are governed by rules on different levels. These rules are often referred to as norms but I will stick to using the word rules for now, as it is more general. The most primitive and fundamental rules are the laws of nature. We cannot change these rules, no matter how much we want. Throughout the ages we have altered our conception of these, but in essence they have not changed. Another kind of rules is government laws. These rules have been created by humans to maintain order. Stories from a past without these have led us to believe that we are better off with them, even though they may constrain our autonomy. Sometimes they are updated to cope with new circumstances or to harmonize with new conceptions of humanity. A third kind of rules is the regulations that exist in certain communities, like housing cooperatives, and in agreements between people. I call these contract regulations. Also these are created to maintain order but are more local and can thus be altered by some procedure that the involved parties have agreed on. For actions not covered by the above types of laws we usually follow etiquette. This is not a matter for any legislative authority but a means to make people's interaction more lean and predictable.

These four categories of laws govern our actions with decreasing strictness. They can all render a punishment if broken but the severity of that punishment and the likelihood that it will be executed varies. Trying to break a law of nature mercilessly leads to retribution. You will always be found guilty and sentenced the appropriate punishment. If you try to fly without being properly fit for the task and therefore jump off a high building you will most likely die. If you jump off a chair you will land on something, commonly the ground or a floor. The effect is possible to predict and disregards whatever your intention was and whatever personal virtues you have.

Breaking a government law leads to punishment if you are caught. The effect is usually reasonably predictable within some range. The legislative authority will (hopefully) scrutinize your case to determine whether you are guilty or not. Your punishment is dependent on the crime that you have

committed and the circumstances under which you did it. Moreover, your intention matters and also your personal background.

Breaking a contract regulation can lead to exclusion from the community or claims for compensation. Your misbehavior may also be overlooked if it was not too serious and you are generally following the regulations. Like with breaking etiquette, the consequences are not always obvious. Since etiquette is not enforced (although sometimes written as postulates in etiquette guide books) it is adapting to the current opinion among people. Also, context is highly relevant. If you misbehave within the family or among your closest friends it may very well be overlooked altogether. If you misbehave in the same way at work, you might destroy your career opportunities. Most of the rules in a written etiquette concern concrete matters like how to greet politely, what to wear at certain occasions and how to dine correctly but also unwritten rules in the society are sometimes included in etiquette.

All of the above systems concern specific actions and apply to behavior that we are familiar with. If we find ourselves in a novel situation there are yet no norms. Neither is it feasible to formulate law, regulation and etiquette so that they would cover all aspects of human life. In fact, the written statements on how to behave are reflecting only a subset of the rules that we abide by. Most of the millions of rules that we follow are not always strict so when we talk about what generally constrains our behavior in society it is better to use the word norms. So if we look at the categorizing backwards, some norms are regulated by etiquette, some by contracts and some by government authorities. The rest are tacit, yet most of us adjust to them without any deeper reflection. As we now have made a common-sense categorizing of the different rules that are constraining our behavior in a society it is time to get to the point, namely to discuss ethics.

I use the term ethics only to describe the study of morality. Some treat the two terms as synonyms, some use them to denote different contexts (e.g. corporate ethics vs. individual morality) but I prefer to separate them as study and practice. It helps me remember that there are always several perspectives when it comes to moral questions. Ethics recognizes the system of individual moralities. I will now attempt to imply that what we call morality is constantly shifting by making an algorithmic excursion into the dawn of humankind.

Let the origin of human life be X. X can be God or some other supernatural being, evolution or a random event, nothing, circular reference or whatever – what it is is not important for the present idea but consequently it is important that it is not important.

Let then the purpose for a living being be Y. This is more interesting. Y is the utter goal that we ought to strive toward with all our actions. This is the reason why morality exists, regardless of how it was created: if everybody is

to get a fair chance to ever be able to know and appreciate Y, our individual aspirations need to be regulated. For Christians Y is to be granted access to heaven. Aristotle and a lot of modern philosophers call Y happiness. To them, happiness is the greatest undefinable state that we vainly try to achieve in all our efforts. Vainly, since it is impossible to specify what ultimate happiness is. Vainly also because we only have our experiences to guide us in how to become happy. We can not know whether we become happy or not by doing something that we yet do not know anything about. Though happiness, for a simple person like myself, sounds like a pretty reasonable Y, I will, out of respect for dissenting ideas, still avoid postulating it. Like X, neither Y is really necessary for the present idea. Let X and Y be variables.

For those who believe that we are a product of evolution, or any other event without consciousness, it is easy to accept that X is not interesting for understanding morality and that Y is merely a variable – something that we cannot say anything about for sure. But then again, for others X is important. If we instead assume that X was an entity with a consciousness, that had a specific purpose with creating life, then X probably had an idea about Y too. Y would then, unless X was particularly vicious, be presented to be attractive to us, so that we would abide by the rules and do our job in the fulfillment of the specific purpose that humanity was created for. If we believe that this is how things are, then we can enter a dogmatic slumber, assured that Y is not variable and thus live happily with following moral principles from the manual that this X gave to us. This also means that Y would not be interesting to investigate, because it has been replaced by the rules. This way of seeing things seems to have worked pretty well for many centuries already. Strangely enough, society has changed anyway. Humanity no longer lives under the same conditions as it did when we were given the rules. Can we really rest assured that we are understanding the rules correctly and can apply them in our current life? Can we be sure that the generations between the original recipients of the manual and us have done a perfect job in transferring the information and adapting it to new conditions? ⁶

Then again, if we on the other hand do not believe that X had any specific purpose for us, then we should neither assume that Y is fixed. If this scenario would be the case, we must accept that Y is something volatile and ungraspable, as in the very first case. It makes it reasonable to believe that Y might even be something that is evolving over time in concord with societies. “Does this imply that the postmodern simplism ‘everything goes’ would be true?” the appalled critic may ask at the threat of such moral relativism; “That all actions are equally moral?”

⁶This is related to the problem of evil found in monotheistic religions. The most common remedy is to acknowledge the free will of humans – a will that can lead to evil. As we live in a society formed by fallible humans, even the religious person should be interested in inspecting the moral laws that we abide by.

Maybe, if the word moral had an absolute connotation. But not all actions would be accepted by the society we live in, even if they are not regulated. We have been socialized to cherish certain norms as moral values and are therefore, consciously or not, disagreeing with actions that do not conform to these. We do not like when people talk too loud in public. We find ourselves honking the horn in honest rage at drivers who are making mistakes in traffic and we detest old men who drool at the sight of young women. We have internalized the norms of our society and thus connected strong emotions to these. We call them moral values. A person's morality, i.e. her set of internalized norms, is closely intertwined with, if not altogether, her identity. Indignation is a beautiful word to describe the feeling that we get when someone steps over the moral boundaries that we unconsciously have defined for ourselves and for our peers. However, I believe that the key to understanding morality is to not so much to be found in scrutinizing the result itself but instead to focus on what created it. People do not generally ponder much over whether they are moral relativists or absolutists, deontologists or consequentialists. Still they lustily participate in making value judgments about others.

Let us play a game and assume that humanity is in fact inherently striving toward Y; that humanity as a collective has a shared but unconscious knowledge about what actually is the purpose of us being here. Assume it is programmed in our genes to survive as a species. It would mean that the morality we express and exercise would be what is showing us the right track towards that, which was happiness for Aristotle. It would also mean that we could use current moral principles to induce the current Y; by observing how people express themselves we can draw conclusions about what Y presently is. This would imply that e.g. immoral comic books, video violence, explicit lyrics, liberal views on sexuality and video games are just manifestations of a shifting morality; that old norms need to be updated. If we assume as an axiom that morality is designed to be purposeful for mankind, then there must be a reason for such a shift of morality. The possible reasons would then be that either Y has shifted or that the current situation requires new mores to realign toward Y or that the norms can be made laxer and still serve their purpose. In any case the morality that we can observe is just descriptive for the direction in which we are moving but can still not say anything about the future.

The point with the absurdities above has partly been to investigate if different perspectives on human evolution affect how moral questions should be treated but also to suggest that the actions and choices that we consider to be moral may not necessarily be absolute. More reasonable is to think about morality as an inert mass that everybody are pulling in various directions. One person has little influence on the direction of this mass but if many

people pull it in approximately the same direction, then our idea of morality will move in line with the resultant.

Some libertarian lawyers oppose legislation based on moral grounds, not because they oppose the morality but only because they believe that laws cannot be based on anything except violations against human rights (e.g. Block et al., 2000; Schultz, 2010). It makes some sense and could paradoxically increase the desire to act morally (see below). Morality can only come in question when a choice situation is at hand. Thus it seems reasonable to claim that the less regulated a society is, the more important ethics become. In a totalitarian system the individuals are “relieved” from personal responsibility for certain decisions. The essence of control is to eliminate choice. In our past, religion has been very successful in this aspect and has worked well to establish a somewhat stable system⁷. Foucault targets this when he analyzes correctional institutions with reference to how the Church used morality to internalize control (Foucault, 1987). Nevertheless, in the current society old dogmas are being rightfully questioned. The libertarian movement in the west is successfully raising doubt about various topics that we once considered to be matters of fact (e.g. Block et al., 2000). At this point we have to be fair and admit that a lot of conditions in our society still are true only by assumption. In order to supply mutually beneficiary relationships, every system needs to maintain constraints on its agents. Law, contracts and etiquette are such constraints. Morality is another. But since morality has no formal power it needs some other mechanism to influence people. That leads us to moralizing.

On Moralizing

Moralizing is the activity that corresponds to morality⁸. This is something that we engage in joyfully throughout our short lives. Whenever we make a value statement about other peoples actions or choices we are participating in this activity. Nagging, agreeing, objecting, laughing, etc. are all different ways to express moral values. Moralizing is a natural behavior for a social animal inhabiting a social system. Most likely 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. Moralizing is simply a way to communicate, negotiate and sustain a complex system of rules. For the reader who prefers to view it from a system

⁷We can never know what would have happened in an alternative system.

⁸I acknowledge that most people have a narrower definition of the term, but as I ascribe moral value to any choice situation and not only those covered by normative ethics, consequently I attribute the same scope to moralizing. This generalized view is purposeful for joining conflicting value systems.

perspective, the goal of societies is to be stable enough to guarantee further prosperity. It seems like Kant (1785/2010) 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 want to call a moral agent. He realized that every rational being in a system have 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, 1785/2010). To prevent morality from following the wrong path (i.e. leading to an unstable system) he identified the categorical imperative as the basis of morality: "So act as if thy maxim were to serve likewise as the universal law (of all rational beings)" (ibid.). This leads to the inclusion of action in the notion of moralizing. A role model is a moralizing agent.

In our current society the role of moral agent is not confined to individuals that interact directly with each other but also includes all sorts of institutions and unidirectional channels: governmental as well as private. According to Gustafsson (1997) media is one of the most efficient agents. He 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 (this description could be also applied to much of social media). He is troubled with the oversimplification and exaggeration that results from several different megaphones competing for attention. It is an interesting topic, especially if new social media is included, but I will leave that discussion out from here. What is important to realize is that we are constantly involved in exchanging moral imperatives with others and that most of us receive more moral judgments than we make. Another interesting aspect of moralizing is that we, when we moralize over others, implicitly also moralize over ourselves. I will qualify this below.

On Morality

A chapter on ethics would not be complete without mentioning Immanuel Kant, a philosopher that recurrently becomes influential in our western ethics. Many interpreters see him as a stern moralist who is imposing duties on us. I agree that he set up hard prerequisites for what conduct to consider moral but disagree on his sternness. It is true that he, based on his own moral conviction, gives some very clear examples of what is good and what is not, but in fact he is prepared to overlook exceptions as long as it is perfectly understood that those exceptions cannot be made universal laws. With the conception of morality as a stabilizing power in society and the impact of moralizing in mind, it is important that we do not even attempt to claim

moral worth of such exceptions. If morality – the result and the source of moralizing – is compromised, the society resting on it will also be.

Recognizing this threat, Kant (1785/2010) writes more about individual morality than about morality as a system in *Fundamental principles of the metaphysics of morals*. Indeed he has the functioning of a general system in mind when he lays out the basic pattern for how individuals ought to behave. Nevertheless, the main focus is to prescribe an ideal disposition for individuals in order for such a general system to function. He is impressively exhaustive, although somewhat difficult to read. Whenever I refer to Kant, I will therefore account for my interpretations. The purpose is to show that a so-called deontological (duty based) perspective on morality does not prescribe following authoritative commands imposed from outside, but demands moral investigation as much as consequence based ethics does.

"Finally, there is an imperative which commands a certain conduct immediately, without having as its condition any other purpose to be attained by it. This imperative is categorical. It concerns not the matter of the action, or its intended result, but its form and the principle of which it is itself a result; and what is essentially good in it consists in the mental disposition, let the consequence be what it may. This imperative may be called that of morality." (Kant, 1785/2010)

In my opinion, this passage sums up the whole basis for Kant's reasoning. First of all it shows that his reasoning is deontological; actions should be judged independent from any consequences. Second, he defines categorical imperatives as moral incentives that we treat as ends in themselves. These can be observed in psychological research (remember protected values from the introduction: Baron and Spranca, 1997) but he reached to the conclusion only by reasoning. I will try to make clear how I interpret it.

The first premise, that skills in any art or craft are always means to achieve some end, should be reasonably clear. We do not need skills for their own sake but to be successful in some action. The second premise is that what I call Y above (Kant uses the word happiness) is an end that cannot be determined as such. Although it would make sense to say "I want to be happy", the decision is not really about choosing happiness instead of something else, but about preferring a mean that is believed to lead to happiness. But since we are referred to our sensations and experience to determine how to reach it, we cannot know for sure what to choose. Kant calls the skill of acquiring happiness prudence and determines from the first premise that this is a mean to an end. Moreover, if prudence would derive from our senses, it would mean that we are dependent on conditions in the outside world to achieve it, which would inevitably be hypothetical. Therefore it must be assumed that we know without qualification what makes us happy, which means that prudence must be derived from pure reason. As reason exists in the world of our intellects, it is commanded only

by the rules of our mental disposition. This means that the only rule that is an end in itself, i.e. that is categorical, is the way we choose actions when we use our free will. The categorical imperative is thus not something that we choose to follow or not: If we know what is good, then we will do it, regardless of any consequences. However, since our reasoning is not clearly disconnected from our inclinations and other constraining influences, we can not know for sure that we really are acting according to the categorical imperative and not for some personal benefit. We thus need to take any possible precaution to counter biases in our reasoning and we cannot rely on any authority to give us the answer:

“Imitation finds no place at all in morality, and examples serve only for encouragement, i.e., they put beyond doubt the feasibility of what the law commands, they make visible that which the practical rule expresses more generally, but they can never authorize us to set aside the true original which lies in reason and to guide ourselves by examples.” (ibid.)

This last quote stresses that, according to Kant, guidelines, examples and policies do serve a purpose to increase awareness of moral problems but can never replace the individual's responsibility to reason about the specific circumstances; the only right way to handle a moral problem is by using autonomous reasoning. Furthermore, it does not assume any specific content for the conception of morality.

Cognitive Dissonance

One important implication from Kant's description of morality is that we need to act according to our morality in order to be moral. It sounds like a truism but this following quote illustrates that he is more likely referring to a psychological consequence of action.

“[...Man] *conceives actions as possible to him, nay, even as necessary* which can only be done by disregarding all desires and sensible inclinations. [...] [It is only in the reality where pure reason alone independent of governing sensations], as an intelligence, that he is his proper self (being as man only the appearance of himself), *those laws apply to him directly and categorically, so that the incitements of inclinations and appetites* (in other words the whole nature of the world of sense) *cannot impair the laws of his volition* as an intelligence. Nay, he does not even hold himself responsible for the former or ascribe them to his proper self, i.e., his will: *he only ascribes to his will any indulgence which he might yield them if he allowed them to influence his maxims* to the prejudice of the rational laws of the will.” (Stresses added)

The implication from this observation is very similar to what Festinger formulated in his influential theory of cognitive dissonance (Brown, 2000),

namely that if, when we are not overpowered by our inclinations, we consciously do something against our will, then we will adjust our will to harmonize with this action. The impact of this cognitive mechanism is thorough. It explains why we habitually internalize moral rules, even if we have not realized their existence or understood their purpose. It also means that if we pose constraints on others for no personal gain (when we do not satisfy our own inclinations), we will also pose the same constraints on ourselves. This is the power of moralizing. A decision maker who is deliberating over moral problems will thus be involved in self-moralizing.

Decision Making Strategies

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 personal interest in the conclusion, and have to reach it in reasonable time, we probably need some 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 the 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.

Psychological evidence and philosophical reasoning lead to the same conclusion, that there are essentially two main processes involved in decision making. The simplest of these is what some psychologists call intuitive judgment, associative reasoning or just System I reasoning, and the more complex one is sometimes called rule-based or System II reasoning (Sloman, 2002; Kahneman, 2002; Sunstein, 2005). Based on the philosophy of Kant and the psychological inquiries of Piaget (1932) I prefer to use the more general terms heteronomy and autonomy respectively. I am sure that not everybody would agree that the meaning of these different notions is identical but for the current application I only need to sketch roughly two types of tendencies in problem solving. These are not mutually excluding and most of the time we involve both in our reasoning so it is not necessary to create an unanimous definition but only to establish the awareness of them.

Heteronomy

The notion associative reasoning could, in my opinion, be a bit misleading, since heteronomy is not about liberated associations but about fixated associations. Heteronomy means that the thinking is constrained by previous knowledge and authorities, rules, biases and heuristics. It includes the automatic, knee-jerk responses to stimuli that relieve us from mental computation and thus constitute a very economic strategy. In most situations heteronomy leads to perfect conclusions but when we are dealing with non-prototypical situations – situations in which we cannot readily apply experiential knowledge – this mental shortcut can misfire and create erring decisions. However, it is important to keep in mind that these heuristics and biases are purposeful in our daily lives (e.g. De Martino et al., 2006). In order to be able to cope with the complexity created by a multitude of interactive every-day events, it is necessary for us to make predictions and assumptions about the likely outcomes of these. In most cases they serve us well and only in certain situations 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, 2002/2010). A recent study on a gender stereotype bias illustrates that it is not only a study of academic relevance. (Ragatz and Russell, 2010). In this case people's judgment of how severe the punishment for a crime-of-passion should be was affected by something other than the crime itself. For the exact same crime scenario, with only sexes and sexual orientations replaced, heterosexual female felons were sentenced to significantly shorter prison terms than either heterosexual men and homosexual men and women. I refrain from discussing details about this particular case, to just conclude that we constantly operate under biases and that it is important to deploy strategies to counter these. Table 1 gives short summaries of a sample of well-known relevant heuristics biases that I will revisit later when I describe the implementation of these.

Autonomy

I suspect that also the word autonomy often gives misleading associations. Computer scientists might imagine a robot that is making its own decisions based on some algorithm. Cognitive psychologists might associate to the autonomous stage in skill acquisition (see e.g. Anderson, 2010). Both of these attributions falsely imply that autonomy has something to do with automation. Also the alternative term rule-based reasoning leads in this direction. I stress that this is the exact opposite interpretation. The above

Table 1. Heuristics and Biases, sampled from Kahneman (2002/2010), Brown (2000) and Anderson (2010).

| | |
|-------------------------------|---|
| Framing effect | The way a choice situation is described affects how we judge the outcome. "Team A defeated Team B" and "Team B lost to Team A" give the same information but render different associations. |
| Anchoring bias | The tendency to rely overly on one piece of information. Related to the primacy effect, which means that we have stronger memory for early events than for subsequent ones |
| Base rate neglect | Underlying statistical probabilities are neglected when making a decision where specifics are present. |
| Choice supportive bias | Presumably a strategy to reduce cognitive dissonance in the conflict between "I preferred X" but "I committed to Y". Therefore "I" search for evidence that upgrade Y and downgrade X. |
| Confirmation bias | Selective search for evidence that confirm preconceptions and reinterpretation of evidence that does not support these. Similar to choice supportive bias. |
| Contrast effect | Can cause a misrepresented evaluation when alternatives are contrasted with other alternatives instead of being evaluated on their own. |
| Maslow's hammer | The tendency to view problems as solvable with available methods. "If all you have is a hammer, then everything looks like a nail." |
| Einstellung effect | If we manage to solve a set of problems repeatedly with the same method, we are prone to use the same method on the next problem. |
| Representativeness | A preference for matching objects into categories that are specific and fitting rather than general categories that are more probable. |
| Fundamental attribution error | A tendency to attribute personality-based explanations rather than situation-based to explain (mis)behavior. |
| In-group favorability bias | A tendency to benefit people perceived as more like oneself. |

processes regard heteronomy. The autonomy of the individual is her ability to reason as a free agent. The mental process for autonomy is pure reasoning, regardless of any authority or inclination.

Jean Piaget (1932) studied children of different ages to derive conclusions about mechanisms of the moral development of an individual. The main contribution from Piaget 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, regarded simple games played by children, where the “spirit of the game” was the very reason for setting up the rules, but not until late in their development they realized this. The generalization of this is that moral rules create order in communities and that autonomy replaces conformity as we mature. 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. Moral maturity, i.e. autonomy, is however 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. For decision makers it is crucial to have a developed ability to reason. The self-doubt that comes along with responsibility is both an inevitable symptom from autonomy and a strong motive to further engage this ability.

Ethical competence

Lawrence Kohlberg (1984) was a pioneer in the classification of ethical competence. In his tests, the respondents were to give explanations for why they considered one option better than another in a situation with ethical implications. Through judging and classifying their answers he defined evolutionary stages of moral development. The advantage with this method of Kohlberg's is that it focuses on the argumentation for a certain choice more than the righteousness of it, but the one crucial disadvantage is that he relied on certain principles to benchmark the respondents' explanations. Thus his own opinions about right and wrong were reflecting the judgment of the answers. To avoid this type of subjectivity it is desirable to avoid classification and bring the definition of ethical competence away from judging the normative contents of an individual's choices, towards a focus on the process of ethical decision making.

Nevertheless, 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. However, this raises a lot more questions than it solves of which perhaps the most

important is how to determine 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 become an expression that could only be used when there is a lack of it. No, the actual judgment of morality should be left to the individual facing the moral problem. To have ethical competence can only be a preparedness to handle moral problems. 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.” (p. 167)

The common conception of a moral problem is one where two principles are heads up with equal strength. Most hypothetical examples 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. All relevant information should be gathered, all stakeholders and their interests taken in 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 above. Autonomy is the necessary attitude to achieve true ethical competence: Not as the ability to always act according to guidelines; not as the ability to act in a manner that is consistent with the most number or best of philosophies; but as a suitable problem-solving and decision-making attitude toward moral problems.

Procedural Ethics

In the late nineties Walter Maner did a survey of 60 methods designed to treat ethical questions in a systematic, procedural way (Maner, 2002). To give a brief overview of the work that has been done in the research area of ethical decision making, I will revisit this key note paper for AICE99 International Computer Ethics Conference, subsequently published also as a journal article: *“Heuristic Methods for Computer Ethics”*. In the paper twelve out of the 60 methods are presented as a representation of different

approaches. It should be stressed 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. Despite the research area, only one of the methods was implemented as computer software at the time of publication. The most valuable contribution of the paper is the attempt to synthesize these different strategies into a twelve-stage process (Table 2) 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 explaining the reasons why certain stages are chosen while others are not. Below I will detail why I consider especially the calculating stage to be unsuitable in an ethical procedure.

As a pedantic side note it can be questioned whether stratification in stages is necessarily the best approach for achieving an optimal result. It seems like thinking in stages has been influential on many “computer ethicists,” probably deriving from early methods for project management. As iterative processes have gained popularity within software development, also the ethical procedures have become more iterative.

On Calculating

Ethics in practice is not just a matter for philosophers. It comes along with gains and losses, all depending on how a situation is handled. It is however important to remember that economical constraints should not come in question before the ethical analysis is satisfactorily prepared. Involving profits and debits in an early stage will block the ability to rational reasoning (Sunstein, 2005).

In “Ethics and decision” Brans (2002) is suggesting an adaptation of multi-criteria calculation method as a way to incorporate subjective and ethical aspects into maxima calculations. In short the idea builds on assigning a linear distribution of weight to vectors, representing different options, in a k -dimensional space (where k is the number of criteria) and then projecting these vectors onto the particular two-dimensional plane that is “preserving the highest percentage of global information” (ibid. p. 349). The distributions of weight will result in elliptical projections of ethically reasonable decisions on the plane, centered on the mean value. This kind of

Table 2. Stages in ethical decision making systems/tools (Maner, 2002).

| | | |
|----|-----------------------|---|
| 1 | The Preparing Stage | Cultivation of moral awareness and acquiring of ethical competence |
| 2 | The Inspecting Stage | Definition of the problem, gathering of facts, identification of stakeholders, relationships, etc. |
| 3 | The Elucidating Stage | Identification and development of facts and presumptions, identification of key issues and framing of problem |
| 4 | The Ascribing Stage | Specification of the values, interests, principles, biases etc., that are the driving forces for a possible conflict |
| 5 | The Optioning Stage | Brainstorming to develop possible alternative actions for participants to solve the problem followed by a screening to eliminate impossible options |
| 6 | The Predicting Stage | Prediction of potential consequences from the different actions identified at the previous stage |
| 7 | The Focusing Stage | Choosing of a set of stakeholders, values or issues to consider more in detail in order to identify the core ethical issue |
| 8 | The Calculating Stage | Quantification and weighing of risks, costs, likelihoods, etc. |
| 9 | The Applying Stage | Deliberation over the gathered information, possibly application of theories and/or weighing of values and arguments for and against options |
| 10 | The Selecting Stage | The choice and common-sense verification of an option possibly leading to a reiteration |
| 11 | The Acting Stage | Planning and carrying through with the decision, development of indicators to assess the consequences of the decision |
| 12 | The Reflecting Stage | Monitoring the implementation of the decision and learning from errors if any, possibly formulating a policy |

blunt description is probably not fair enough to give the reader a chance to the numerical method as such, but it is not really necessary. The reason why I even bring up this approach towards incorporating ethics into multi-criteria calculation is to illustrate that ethics can be interpreted in different ways. It is

difficult to imagine how applying mathematics directly on ethical principles can be done. How is such an ethical framework? Is it possible and judge realistic to assign weights, i.e. ranks of importance, to ethical principles and values? Brans tries to tackle the uncertainty about an absolute value by introducing into the calculus a distribution, within which the weight is considered to vary. If we for a moment accept such a concept, then we immediately stumble over the next question mark. How are different ethical considerations compared? Can different principles really be pitted against each other? In common practice, multi-criteria calculations are used to analyze multivariable decision problems where criteria naturally become conflicting due to assertions of incompatible numerical values in the calculus (e.g. manufacturing cost could become conflicted with quality). These conflicts are exploited to determine a course of action. The numerical, or at least comparable, character of criteria allows for this type of seeming incompatibility to occur, but for ethical values and principles we cannot credibly assign e.g. investment and maintenance costs to create such a basis for comparison. An ethical principle, as well as a stakeholder interest, is an end in its own, in most cases uncorrelated with other principles, and as such it refuses to become subject to ranking of importance.

From a psychological perspective, decision making is a process of obtaining information. From a mathematical it is a matter of computing available criteria. To join these two approaches, attention should be focused on the assertion that ethically derived solutions in fact are comparable, like any other solutions, even though the ethical principles behind them are not (Wenstøp and Myrmel, 2006). Instead of being the subject of calculations, ethical considerations should contribute in the selection of criteria for calculations. This is where an ethical analysis serves a purpose. The ethical aspects of decision making should not as much be the concern about which of two conflicting principles to offend, as it should be the awareness and recognition of specific humanitarian and ecological values in the process of accounting. Ethics as principles do not need to be quantifiable in order to quantify ethically. If included at all, calculating should be a part of the applying stage and not a foundation for it.

Related work

Many of the tools that address ethical concerns make either an implicit or explicit assumption that using ethical theories can help us make ethically better decisions. This presupposes that the ethically correct way to manage matters is readily identifiable which might be an overly optimistic assumption.

Various ethical support systems have targeted the concern of identifying relevant information in different ways. In Paramedic Ethics (Collins and Miller, 1992) focus is put on the obligations and responsibilities of the decision maker. Based on these, the user is establishing relationships between stakeholders and then identifying considerations for the different opportunities and vulnerabilities that come from alternative solutions. Finally a negotiated social contract alternative is evaluated as a possible compromise solution. In SoDIS (Gotterbarn, 2002) 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. 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. It should be noted that the first two of these systems 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.

Value Sensitive Design, proposed by Friedman, Kahn and Borning (2008), is described by the authors as an interactional theory that takes in account all imaginable values that various philosophies have found to exist. However, those approaches do not focus exclusively on what psychological theory and research describe as the basis of competent ethical problem solving and decision making, namely the readiness for autonomous reasoning (Kohlberg, 1985; Piaget, 1932). The need is rather for methods that promote autonomy and prevent us from using heteronomy when it is not appropriate. All above methods are excellent to systematize, organize and guide the user in concrete moral issues, but since they to different degrees urge and lead the user to moral philosophical contemplation, there is a risk that the user gets lost and thus misses the obvious main goal, namely to handle the practical problem. Of course this is the espoused goal of these methods, but as they include analysis of or comparison to different normative moral theories and in some cases even propose moral solutions (for example Davidrajuh, 2008), they shift the attention from the urgent to the abstract. They do not focus on, nor address, the different impact of heteronomy and autonomy which means that they cannot secure that decisions are not founded solely on heuristics, biases and authoritative attitudes, i.e. heteronomy.

EthXpert

EthXpert⁹, as described in all included papers, is a tool intended to complement codes of conduct and guidelines and help decision makers in the process of establishing a conception, as complete as possible, about a problem with moral implications. It builds on the assumption that ethical competence is equivalent to a well-functioning problem-solving strategy (Kavathatzopoulos, 2003, 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 it impossible to guide the user by asking questions about previously known reasons for moral problems. Also, since the definition of what constitutes an ethical decision cannot be assumed to be at a fix point, this kind of tool must be designed so that it does not make any assertions of the normative correctness in any decisions or statements. This absence of guidelines might initially seem like a weak strategy but 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. Ethical deliberation should leave a feeling of insecurity and open up for self-criticism. Such a setup delegates the responsibility for a satisfactory analysis completely to the analyst. 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. It is however not the only imaginable application area. Moral problems are generally complex and involve lesser or greater number of stakeholders and interests. Based on trials with the P&P method (Kavathatzopoulos, 2003, 2004; Erlandsson and Kavathatzopoulos, 2005; Rauschmayer et al., 2009) and on the theoretical basis of Kant and Piaget, I conclude that the matrix representation, with stakeholders and interests on one axis and alternative solutions on the other, serves fairly well for promoting a systematic assembly of available information about a problem. In the process, options and interests of each imaginable stakeholder are identified, and the matrix cells are filled with gains and risks. The question that I approach with EthXpert is how this process can be improved by computerization. The P&P method supports a holistic view, yet 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

⁹The software can be downloaded from <http://www.it.uu.se/research/project/ethcomp/ethxpert>

process of inputting the data, as well as from a more configurable visualization of the same. 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, are introduced with EthXpert.

Table 3. Requirements on ethical decision support systems/tools

| | |
|---|---|
| Regarding the user/decision maker | Should not have to know a lot of different ethical theories |
| | Should not have to be skilled in causal and consequential reasoning |
| Regarding the stakeholders that are taken into account | Should not have to share ethical principles, codes, laws or policies |
| | Should not have to share moral values |
| General desirables regarding the tool (adapted from Rick in Kavathatzopoulos, Laaksoharju and Rick, 2007) | Should not require or derive from a predefined set of moral principles and values |
| | Should help the user to systematically solve the moral problem at hand |
| | Should help the user to be unconstrained by moral fixations and authorities |
| | Should help to identify and consider as many relevant values and alternate actions as possible |
| | Should encourage the user to motivate his or her decisions in regard to relevant interests and values |
| | Should help the user to organize and analyze the facts |
| | Should help the user to weigh the relevant values and principles against each other |

To create the best possible conditions for a well-founded decision it is important to allow the user to input any imaginable data, without having to worry about the extent of data muddling the lucidity 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 overanalyzing (Brown, 2000, Chapter 5). Therefore the inputting of data is simplified as much as possible. Still any bit of information that might affect the decision process should be possible to add, and 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 the reasons for it should be noted. EthXpert is designed to trigger further considerations, through association and invitation to fill out blanks. Collins and Miller (1992) agree with the benefits from expanding the problem through an associative process and Paramedic Ethics is in many aspects similar to EthXpert. The main difference is however the more important: in Paramedic, stress is on normative obligations, vulnerabilities and opportunities while in EthXpert focus is on how the interests of different stakeholders relate. This makes the analysis procedures different. Where Collins and Miller urge normative valuing of pros and cons, EthXpert only creates a foundation for thorough autonomous analysis. The assumption for this approach is that an unconstrained and not moralizing approach liberates the decision maker and thus enables her to take into consideration all relevant aspects. Besides the requirement to allow inputting unconstrained amounts of data, the requirements on an ethical decision support tool consist of the items listed in Table 3.

In the following paragraphs I will try to show how these requests are fulfilled. The items that regard shared values and ethical theories are handled implicitly by deploying a non-normative approach. The remaining six are dealt with through supplying a structure to support the understanding of how information about moral problems relate to the stakeholders involved.

To support autonomy, the main requirement on the tool is that it should not be making any decisions and not even supporting any specific solutions: it should not be elevated to an authority. The tool should not even give any directions about the correctness of any conclusion. This will force the user to analyze the problem very carefully. The sole intention should be to help the user to organize and structure a problem at hand. At the same time the problem should not be narrowed down, thus risking oversimplification, but instead be expanded and widened so that the user can appreciate the full impact of a decision. Further it should help the user to 1) block heteronomy and support autonomy, 2) organize interrelationships and data in a systematic way and, 3) present the complexity of the issue in a comprehensive way which means to provide easy access to all data.

It goes without saying that it is very hard to decide in moral problem situations – it is in the nature of the topic. The conflicting principles and values behind stakeholder interests will all seem too important to trade off

and the outcome of any realistic option will appear to have undesired features. This is the common perception of an ethical problem.

However, as stated above, in real life many moral problems occur instead from the lack of information or misinterpretation of responsibilities. The intentions might be good but if sensitive information is missing a wrong decision may be taken. Therefore it is important to allow the person 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 but the problem with massive amounts data is however apparent; the chance to make use of it decreases with the amount. Many of the approaches presented in earlier work (Maner, 2002) suggest different strategies to eliminate matters that are not relevant for the problem. The impending risk with elimination is to lose important aspects. A better approach is to let the decision maker be selective when it comes to analyzing the data.

EthXpert supplies configurable representations, where only the data associated to a specific part of the problem is viewed (see below). 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 of course be programmed to choose considerations based on an algorithm for ethical analysis, but as Maner concludes before presenting and refuting an algorithmic interpretation of ethical problem solving: “Ethical problems are too complex and too fluid to solve algorithmically in human time” (Maner, 2002, p 340). Automating the definition of what is relevant is a sure way of elevating heteronomy. The user may start relying on the tool to make the right choices. 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.

Bias Prevention

To support the decision-making process, EthXpert would optimally either block biases or make these obvious to the user. By provoking the user to expand the problem scope and by systematizing the data, EthXpert unlocks constraints. To counter biases and to block any tendency to involve emotional or prejudice, the analysis is divided into two main parts, one where relationship considerations are stated and one where different options and scenarios are considered. In the latter, preference for certain stakeholders or certain decision should not be ruling the considerations. To approach this, the relationship considerations are presented with affecting stakeholders de-

identified. Table 4 brings up the previously mentioned biases and how they may be prevented.

Table 4. Biases addressed in EthXpert

| Bias | Countermeasure |
|---|--|
| Anchoring bias, choice-supportive bias and confirmation bias. | With a systematic scrutiny that leaves the optioning to the last steps the risk for unconsciously omitting relevant information decreases. 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. |
| Framing effect, contrast effect, einstellung effect, Maslow's hammer and base-rate neglect. | These are all related to the way information is perceived. The three different views in the tool all show different aspects of the situation and can somewhat serve to counter framing. Further, the absence of weighting counters a bias to estimate importance of stakeholders. The possibility to easily create compromise options could help counter the fixation on certain alternatives. |
| Representativeness, fundamental attribution error and in-group favorability bias. | These are all related to biases for specific stakeholders. The countermeasure is to first let the user take the perspective of all stakeholders and subsequently depersonalize these in the next step. When analyzing the considerations, the identities of stakeholders are obscured. This will leave only the statements as basis for a decision. |
| Underestimation of uncertainty and undue limitation of problem. | With an expanded view of the problem, the risk for underestimation and oversimplification will most likely be reduced. |

States of Ethical Analysis

EthXpert is a tool to aid primarily the *analyzing* part of the decision process. From Maner's above set of stages (Table 2) it will cover stages 2 to 6 – inspecting, elucidating, ascribing, optioning and predicting. The reasons for omitting stage 8 – calculating – has already been established above. Stage 7 – focusing – is 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. 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 (Paper IV).

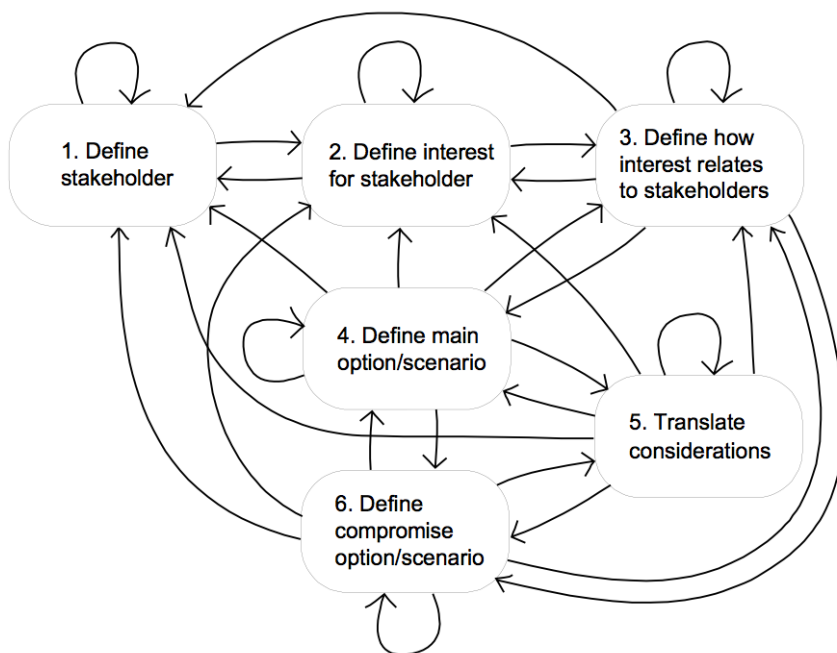


Illustration 2. The six states in the EthXpert procedure.

To stress that the procedure in EthXpert is not hierarchical I have come to adopt term state in place of stage. Three details are especially interesting to note in the state machine representation of the ethical procedure (Illustration 2). The first is how the flow from state 1 to state 3 makes it invalid to enter state 4 without passing through state 3, the second that it is possible to reach all other states from state 4 to state 6. This is a deliberate strategy to ensure focus on elaborating considerations and stakeholder relations. The final peculiarity is the lack of a final state, which is a violation against fundamental programming practice, implying that an optimal state can never be reached. This is in accordance with Boulding's (1966) widening of 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 believe that the analysis is completed.

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 (Illustration 3).

The purpose with this simple task is to establish a clear focus on the problem and the factual conditions of it.

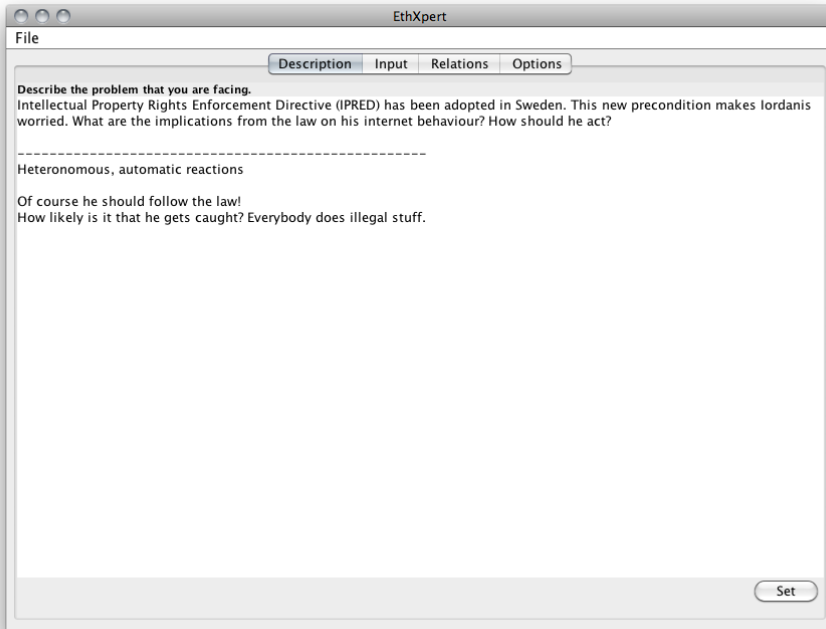


Illustration 3: The view in EthXpert where the user can describe the problem for future reference.

1. Define stakeholder

It is not an obvious task to identify the stakeholders that affect or are affected by a decision. Not even among scholars specialized in stakeholder theory the concept of stakeholders has always been unanimous (Donaldson and Preston, 1995). I agree that the notion of stakeholders can be further elaborated and I should address the work done in this field. A Google search on the term “stakeholder analysis tool” gives over 50.000 results (Google, 2010) but on further inspection, none of the tools seem to have the same explicit focus on how interests affect and are affected by decisions. Most tools in fact seem to aim at identifying and ranking the importance of stakeholders or some other factors (e.g. Moodley et al., 2008). For this ethical analysis process I have chosen to regard everything from society, to organizations and corporations, to individuals as equal stakeholders. Everybody who affect or are affected by, gain or lose on 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 EthXpert, an implicit question will guide the user to increase the problem scope: who is affected by or affecting a specific interest of a stakeholder (Illustration 4)?

2. Define interest for stakeholder

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 user determines for each stakeholder a set of relevant interests, which could very well be unique for it. 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.

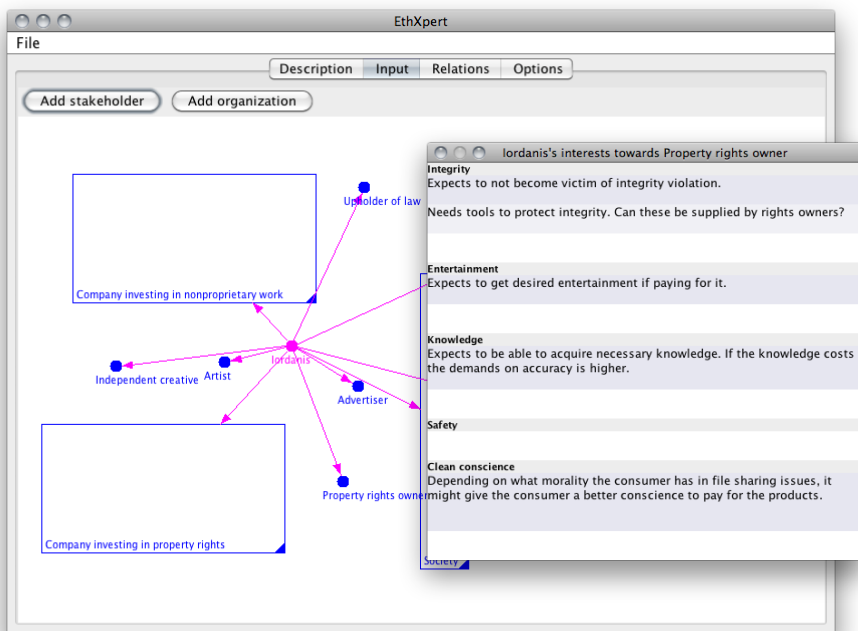


Illustration 4. The view in EthXpert in which stakeholders and interests are defined.

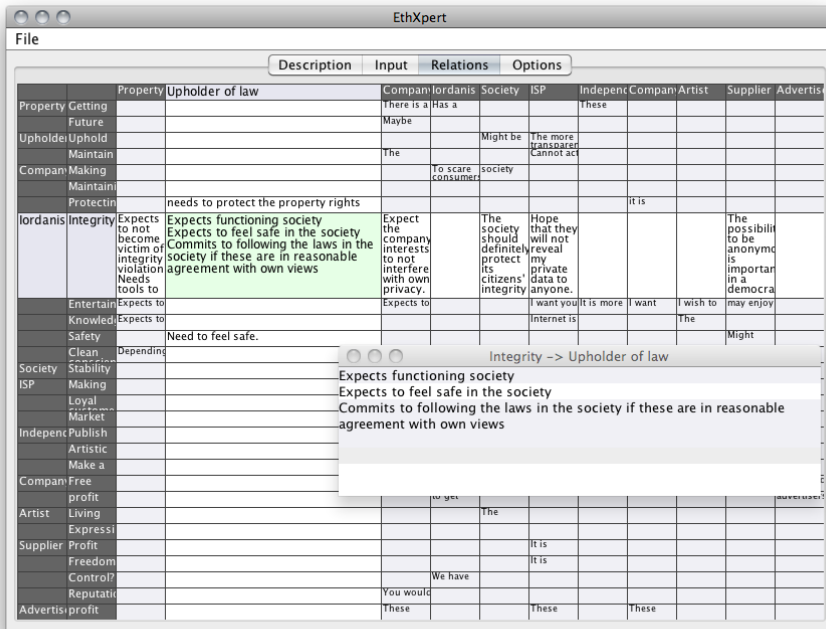


Illustration 5. The view in EthXpert where relationships between interests and stakeholders can be scrutinized.

3. Define how interest relates to stakeholders

The user is urged to try to imagine how the identified interests affect and are affected by other stakeholders (Illustration 5). 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 also helps to track down 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.

4. Define main option/scenario

After all the relationships between stakeholders have been exhaustively analyzed the user can start considering what options there are to handle the problem. 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”.

5. Translate considerations

The considerations from the interest-stakeholder matrix will not be automatically copied to the 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 (Illustration 6).

6. Define compromise option/scenario

To counter problems in the main options, i.e. unacceptable negative effects, compromise scenarios can be spawned from existing options. A compromise

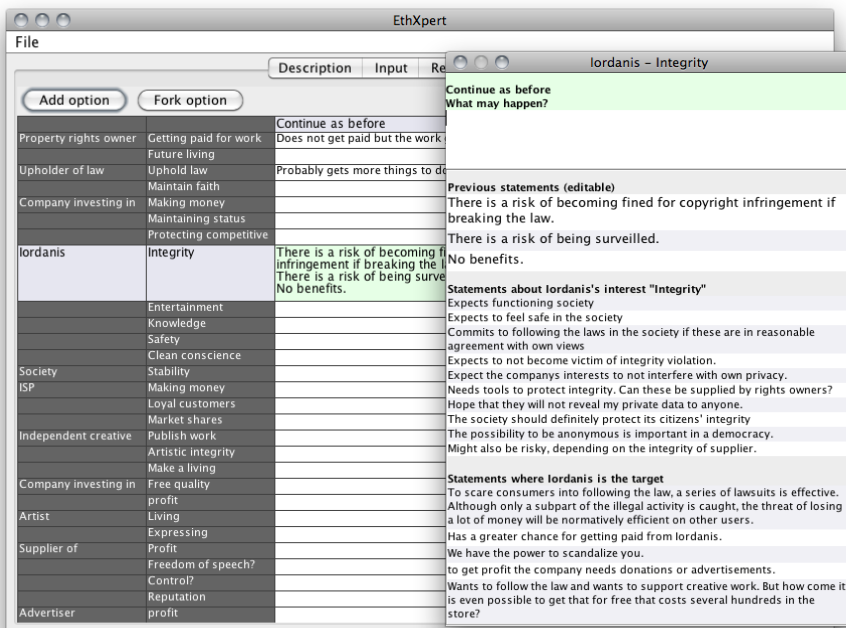


Illustration 6. The view in EthXpert where options/scenarios are evaluated.

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.

Technical description

The novelty of EthXpert lies not in the chosen technical solutions but in the way it supports ethical decision making without making normative claims. The decisions made in the designing of the interface should however still be motivated. The most striking feature of the tool is the way it represents tabular data. Instead of supplying the user with a viewport – a rectangular “window” through which a portion of a table can be studied (i.e. if it is bigger than the viewport) – it condenses the table to fit in one screen with only a focused part of the table fully sized. This technique is often referred to as a fisheye lens and is part of the focus+context design paradigm. The incentive for this way of displaying data is the assumption that contextual information is important for the sense making of specific information. The fisheye technique is not new. It was suggested already in 1982 and tried out during the beginning of the nineties but not until TableLens was presented in 1994 did it gain wider recognition (Rao and Card, 1994). The idea for TableLens was to display very long lists¹⁰ of mainly numerical data in one single display (about 64,800 cells on a 19” display) by reducing unfocused information to a graphical representation, showing the relative magnitude of the data. The main benefit in TableLens is that it takes advantage of the human's excellent cognitive ability for pattern recognition.

Since then, several other solutions have been suggested (e.g. Klein et al., 2002) but the work most relevant to me was done at the beginning of the 21st century. DateLens is a nice calendar application with a fisheye interface (Bederson and Clamage, 2004). The effect here, like in HyperGrid – a zoomable table interface (Jetter et al., 2006), is that focusing, i.e. “clicking”, on one data cell reveals additional specific information while the spatial relation to other cells is maintained. The point, according to Jetter et al. is to avoid that the user gets lost in “table space”, i.e. loses track of where in a big table the sought-for information can be found. The motive makes sense in the case of HyperGrid, which is intended for large databases of textual information, but the benefit is less obvious in DateLens. The technique increases usability mainly when it is important to keep contextual

¹⁰A great number of rows but few columns

information visible. In a calendar application it is not apparent why this is necessary. However, it should be mentioned that DateLens was targeted at PDAs with small screens but the question remains whether contextual overview is necessary in a calendar. The usability testing performed with DateLens reveals that indeed, the fisheye technique was most effective for calendar tasks that depended on both contextual and specific information, e.g. to find out which two weeks period that has the least number of scheduled appointments. In a calendar, this kind of tasks is probably fairly rare, while in EthXpert this is the predominant approach. Another distinction is that EthXpert distorts the table to fit in one screen by shrinking and clipping the text in unfocused cells. This is more similar to TableLens although it poses an undesirable constraint on the possible amount of data. The currently evaluated version allows a maximum number of about 70 interests and 20 stakeholders on a 1440x900 pixels display and even this number reduces the usefulness of contextual information considerably due to cells being too small to display.

In order to make the analysis work as lean as possible, the number of required interactions are kept at a minimum for the most common actions. E.g. when clicking on a cell, the dialog for inputting information is immediately opened and the input field is focused. When exiting the field the information is automatically saved unless the input is explicitly aborted. Addition of stakeholders and interests automatically includes these in the analysis. Navigation is possible with either mouse or keyboard or both simultaneously. The latter technique allows the user to investigate other relationships while working with one. Mouse tooltip labels are used to show the content of cells that are too small to display everything.

Methodology and Methods

When I have talked about my research with my colleagues, I have often heard that it resides on the outskirts of the HCI research area. Ethics is a concern for philosophers and computerized decision support systems usually give associations to knowledge management, which is usually investigated within information or business science. It is indeed not the mainstream topic in HCI. Still I claim that what I do is perfectly fit within the scope of HCI. My research interest is not to deliberate over ethical issues resulting from the use of ICT, nor is it to study organizational processes in the treatment of ethical problems. I design tools that assist people other than myself in their handling of ethical problems. In order to be successful at this I need knowledge about the domain where the tools are to be used. This is good HCI practice.

Thus, I need to know what philosophers think about ethics. I need to know what psychologists have concluded about moral reasoning and I need to know how ethical problems are handled in current practice. In essence, this corresponds to the common use of ethnography, although lacking a name and prescripts for how to attain the situated knowledge.

Most of this thesis covers theory on ethical competence (described further down). The method for this part of the work has been to analyze how philosophers approach situations with ethical implications, in order to derive a general method of handling moral problems. The tool that this thesis covers is based on this analysis. The procedure in the tool is thus deduced from theory. To support the process of philosophizing, the tool is designed to block the natural tendency to be biased and to lead the user to expand the viewport that is constraining the understanding of the situation. The tool has subsequently been tested in comparative studies where participants approached various ethical problems. These studies have rendered both qualitative and quantitative data, which has been used to further improve the tool inductively in accordance with theory.

Usability of EthXpert

Papers II and III describes studies on the usability, in terms of efficiency, effectiveness and satisfaction, of EthXpert (further described below). In the

studies, the scope of consideration for options, stakeholders and stakeholder interests were assessed by counting the number of items in each of the categories. The question to be answered was whether a computerized tool can support the problem understanding better than a corresponding paper-and-pen based tool (P&P).

Study 1

The first study, presented in paper II, was conducted mainly to investigate the usability of the software EthXpert, both in the sense of how well the application supported the process of critical scrutinizing of ethical problems, and in the more practical sense of how well the participants could utilize the functionality of the software. In the study six groups, comprising between six and seven participants each, all of whom were third or fourth year students at the Department of Information Technology at Uppsala University, evaluated the ethical implications from introducing new IT systems. Prior to the study, the participants received lectures, made assignments and participated in seminars about computer ethics and ethical usability.

In the study the participants were given two tasks. The first task was to make an analysis of the chosen problem using the OLE questionnaire and the P&P method. The second task was to use the first alpha release of EthXpert for the same analysis. After the two tasks the participants were asked to give written feedback on how well EthXpert supported the process of autonomously handling ethical problems as well as on how usable the software was generally.

Study 2

For the second study, presented in paper III, EthXpert was slightly improved, based on the reporting of usability problems from the first study. These changes included bug fixing and interface improvements but the basic structure of the program remained unaltered.

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.

Prior to the task they were given an hour and a half introductory training in how to use the tools. They were then divided into two groups: One with five participants and the other with six participants. Both groups started by doing an OLE analysis after which one group continued with EthXpert for a second analysis, followed by P&P for a third, and the other group used the tools in reverse order. After the analyses they were asked to grade the tools

in a questionnaire (see appendix 2, paper III) and write a 100-300 words evaluation of the tools.

Analysis

The first exploratory study, presented in paper II, gave only tentative and formative results that were not analyzed in detail. The second study, presented in paper III, rendered four types of data.

Quantitative data

- 1) The numbers of identified stakeholders, interests and options from EthXpert and P&P.
- 2) The ratings from the questionnaires.

The significance of the differences in identified stakeholders and interests between the two groups was calculated with Student's t-test. The ratings did not render conclusive results and were used only as indicative. I will revisit them with cross-reference to the qualitative data in this summary.

Qualitative data

- 3) The ethical analyses made by the participants.
- 4) The written summaries supplied qualitative data.

In paper III, the written summaries were categorized into emergent categories. These categories will be elaborated on more in detail in this summary. The participants' analyses were exempted from scrutiny.

Results

This far, two studies of EthXpert have been conducted. The results of these are positive regarding the usefulness of the tool. In both, comparative measures showed that users were able to identify more stakeholders and interests with the assistance of the tool than with the corresponding paper-and-pen-based method (P&P). In neither of the studies did the identification of alternative strategies benefit from using the computer-based tool.

The first study, described in paper II, was an exploratory study where we wanted to get an indication whether the hypothesis, that EthXpert would help users to identify more stakeholders and interests, was reasonable. The results were slightly positive, but the participants reported that they had suffered from bugs and usability problems in the software.

The results from the second study, presented in paper III, are more interesting. The identification of stakeholders and interests benefited significantly from using the computerized tool. Furthermore, the order in which the tools were used played a significant role in how many of these were identified. The group that started using EthXpert identified almost twice as many stakeholders and well more than twice as many interests than the group that started with P&P.

Satisfaction Ratings

The overall results from the questionnaires were mildly positive for both P&P and EthXpert (average score of 0.52 and 0.50 on a scale from -2 to 2 where 0 is neutral). For OLE the results were a bit below neutral (-0.27). Both groups started with the OLE analysis and the fact that they gave the same average score suggests that the group division was unbiased with regard to attitude toward the autonomy approach. Due to the small sample space no significant correlations could be calculated, but there still are some observations worth mentioning. In paper III it is indicated that the rating of P&P was more dependent on the order in which the tools were used than EthXpert. This impression is supported by the numbers of identified stakeholders and interests, as participants in both groups involved a larger scope of the problem in EthXpert than in P&P. The P&P method thus did not give additional help to extend the problem for the participants who started with EthXpert but could perhaps have helped some participants to focus the

attention on a particular subset of stakeholders. One peculiar observation is the covariance between the ratings of EthXpert and P&P within the two groups. This points at a problem in the design of the questionnaire, which reduces the validity of the results.

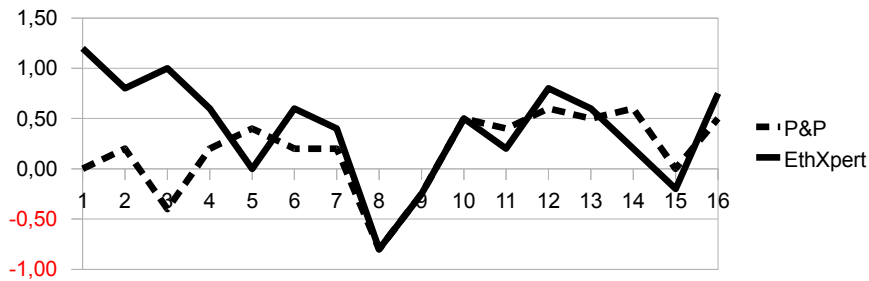


Illustration 7. The ratings of the group that started with EthXpert. Questions 1-16 (appendix 2, paper III) on x-axis and ratings on y-axis (-2 negative – 2 positive).

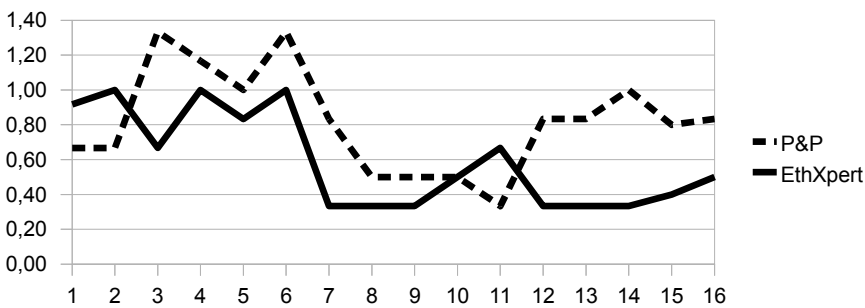


Illustration 8. The ratings of the group that started with P&P. Questions 1-16 (appendix 2, paper III) on x-axis and ratings on y-axis (-2 negative – 2 positive).

Qualitative Analysis

“I thought I would really like ethXpert since I write much faster on a computer than with a pen. I filled out all stake-holders and all interests I could think of. Then I started stating relationships. That's when I started realizing this was not the best method to do it after all. Because I had listed a lot of stakeholders and interests, it took a really long time to fill out all relationships, and I didn't even fill out all of them in the end. Things didn't get

better when I came to the options part. Again, there were so many cells to fill out and so many statements (that I did in the relationship-part) to consider for every risk and possibility that I completely lost the overview and gave up.” (Anonymous participant in Study 2)

The above statement describes the problems that a person may face when using EthXpert for the first time. Indeed the tool is deceptively simple. There are only four tabs and a handful of functions but with these the user can unknowingly create a complex system of relationships that demands attention. While overviewing the graphical stakeholder network or investigating each stakeholder's interests in isolation, it may seem very straightforward and clear, but once you switch tab to the first matrix you start realizing that the system can easily become overwhelming. This is the challenge in ethical decisions, as social consequences tend to have far-reaching impact. Currently, the tool does not give much help to bring order. The written reports of the participants on the strengths and weaknesses in EthXpert are summarized below with additional comments on how they relate to the underlying theory.

Positive

- easy to name stakeholders and helpful to detect interests; – easy to add stakeholders and interests; – very easy to define stakeholder interests and expectations from other stakeholders

In order for the tool to support a flow of associations the addition of stakeholders and interests has to be effortless. It seems like the mechanism is already appreciated by some users but the association process can be further improved by letting users add stakeholders and interests from any of the three views. As can be seen in Illustration 7 and Illustration 8, both groups gave relatively high scores to EthXpert on the first two questions, which were concerning how good the tool was in identifying stakeholders and stakeholder interests.

- best for problems with many stakeholders, interests and outcomes; – helps to cover many aspects of the analysis; – identified more possibilities than in P&P; – good approach – especially for those who prefer stimulus or focus in ethical questions; – allows focus at parts of a problem; – expanding the overall picture; – suitable for people who like to constrain the problem and focus on parts in isolation

These quotes support two important requirements on the tool. First, the tool can help users to expand the problem scope and second, it helps them to focus on parts of a problem.

– less dependent on external factors [i.e. personal traits] than OLE

to put yourself in the position of others is inspiring and gives new insights; – the interest relations adds interesting depth and helps detaching from one's limited point of view

It seems like the process in the tool supports users in taking the perspective of others, which is something desirable but impossible to truly achieve. These statements express subjective opinions but perhaps it is enough that a user gets help to investigate a scenario from someone else's point of view without necessarily consulting this other stakeholder. Such a procedure naturally opens up for strong criticism so it is in the best interest of the decision maker to make an unbiased scrutinizing of the problem. In fact, involving outside stakeholders may actually reduce the feeling of responsibility for the decision process.

– made it very clear how outcomes would affect stakeholder interests

This is naturally the main point of the tool so it is nice to observe that it made an explicit impact on one of the test subjects. At first it may come as a bit of a surprise that ethical situations can be systematized in this

– good overview; – everything visible; – stakeholders were shown in a good way; – very structured

One of the leading requirements in the design of the tool has been to support the overview of the analyzed situation. The matrix views are designed so that all information is concurrently accessible in order to support associations. Cognitively this means that the user has equal access to many sources of stimulus, regardless of where it is positioned in the matrix. Further it helps the awareness of the boundaries of the situation. The ratings support these statements (question 6 in Illustration 7 and Illustration 8).

– quite interesting to use; – user-friendly, – easy to understand; – relatively intuitive; – easy to revise;

From a user experience point of view it is good if a tool is not only helping the user to fulfill a task but also makes the work stimulating. These opinions are however only mildly reflected in the ratings of the software (questions 13 and 14 in Illustration 7 and Illustration 8).

Mixed

- it turned into a combinatorial explosion, fields that needed filling in expanded rapidly

This is an understandable reaction as the process in the tool is seductive. The user starts by assigning interests to stakeholders, perhaps without realizing the effect of each interest initially relating to each of the other stakeholders. If done ambitiously, she will find a huge matrix, begging for content, when she changes the tab. Although perhaps shocking, this is actually a helpful feature. As the growing complexity is initially hidden, the user is less likely to make a premature screening of possible interests.

- started with wrong strategy but would now prefer to use EthXpert for future problems
- would like to see the heteronomy part in EthXpert and also a concluding part and coloring of cells

I have been considering a concluding part where the final decision is vindicated with references to the relevant cells in the matrix. Coloring of cells is also an item high up on the todo list.

- sometimes considerations are similar – would like to be able to link these

This too is a good suggestion that I have been thinking about how to implement so that it is apparent that they are related but without making the interface overloaded with information. It is trivial if the cells that are to be linked are neighboring since they then can be merged, but this setup cannot be guaranteed. The obvious approach is to color code cells that are related in the same color, but this both puts a limit on how many different groups can be created and renders a messy appearance. The positive aspects in this kind of feature are that it would assist the user in both identifying and scrutinizing patterns existing in the data, but on the negative side it could also lead to a misdirected pursuit for manipulating data to fit into neat patterns. I stress that the point of the tool is not to establish a neat interpretation of a situation but to display the full complexity in it.

Negative

- grows fast and becomes impossible to handle; – too much information – hard to select what is important; – more time consuming; – lost the overview due to the high number of stakeholders and interests

- messy setup – did not fit my way of thinking; – messier than P&P – no overview
- difficult to envision every relationship between stakeholder interests and stakeholders
- difficult to focus on problem (as in P&P)

It seems like the main problem perceived with EthXpert is that it does not help the user to make a decision (question 8 in Illustration 7 and Illustration 8). Also from these comments it should become apparent that EthXpert is not a tool that will make decision making easy. It is positive that the users were able to increase the scope of the problem but it is not satisfactory that they in doing so lost control over the problem. Perhaps the tool should help the user to identify aspects of the problem that could be analyzed separately, perhaps there needs to be stronger support for focusing the scope or perhaps the interface should help the user to somehow filter out redundant information. The simple answer that avoids the problem would be that the user needs more training.

- difficult to explain, discuss and solve the problem

In order to understand this comment it is necessary to revisit the analysis of the specific participant. It reveals that the person has failed to perform the analysis except for the first step. On average the users were mildly positive about the usability of the tool (questions 13 and 14 in Illustration 7 and Illustration 8) so this again stresses the importance to make the tool easier to start using.

- not a good overview when deciding options (interest considerations are not directed towards specific stakeholders anymore)

Even though this may be annoying to some users, it is actually a deliberate design decision. In order to minimize biases for or against particular stakeholders, the statements made in the relationship matrix are anonymized in the decision evaluation state.

- a bit tricky to use; – a bit buggy; – same as P&P but slightly more complicated; – took a while to understand how relations worked; – difficult to understand how to start; – requires more education
- the potential of the stakeholder network is not fully utilized

I acknowledge that the system is neither complete nor intuitive in all aspects. Some unnecessary problems are created from the use of unclear terminology

and a lack of instructive guidelines. Others depend on insufficient assistance from the tool. The ratings on questionnaire items concerning usability are however not fully consistent with these comments (questions 12 and 13 in Illustration 7 and Illustration 8). Nevertheless, there are a number of usability problems that should not be present in software from a HCI research department: The font size is fixed and in its smallest form too small for some users, the color theme is fixed and unsuitable for people who need stark contrasts, it is not possible to change language and the terms used may not be understandable for all users, there are no help pages, keyboard accelerators are not following any standard, warning dialogs appear also when they are not necessary and the tool does not remember the user's preferred window setup between sessions.

These problems (and probably several others) are important to take care of if the tool is to be released to the public but as the tool currently is just for illustrating a concept it has not been prioritized.

Discussion

The results are positive but pose interesting challenges. In both studies EthXpert helped the participants to identify further stakeholders and values, which allowed them to include these in the analysis, but then in some cases they felt like they failed to make use of the extensive information. Theoretically this expansion of the problem is exactly what is needed in decision making – it prevents the user from missing important aspects – but the information has to be made usable in practice. Thus, there is room for improvements. In the versions that have been tested there are no features to help the user disregard irrelevant information. The encouraged attitude of adding possibly irrelevant considerations rather than omitting them therefore can result in the user losing control over the analyzed situation. This is just what can be expected from information overflow: The more information that is added, the more the clarity of the problem is obscured. Ironically, this is exactly what the system was intended to prevent. Indeed, the intention has deliberately been to aid a widening of moral problems and it is true that such an approach does not always make the choice easier. Nevertheless, considering the expanding problem scope, the tool needs to supply the user with more mechanisms to suppress information that is not currently relevant and further systematize the information that is interesting. The tool should somehow help highlight topics where further investigation and discussion is needed. This could be either a feature that automatically classifies considerations according to some technical rules, or interface features that let the user herself do the classification. The benefits from the first approach would be that it would release the user from the extra burden of thinking about what taxonomy to apply in the particular case and it could also allow for identification of latent dependencies between different considerations. However, there is also a great risk that the automation would elevate the system somewhat to an authority, in which case the user would not easily want to counter the decisions made by the system.

The advocated autonomy method does not automatically help the user to consider all relevant normative standards or to identify relevant principles for the case at hand. This is intentional. The method is constructed to help the user focus on the core dynamics of a problem – the values and interests of stakeholders and the risks and possibilities in all options. In the users' analyses from the first exploratory study I noticed that the approach seems to

be a bit confusing to some users. Several participants missed apparent ethical conflicts in their chosen cases. For instance, in an analysis of using lie detectors in an insurance company's call center no concern was paid to the possible distress and invasion of privacy when monitoring the calls of customers. Neither was the problem of determining how to deal with possible false alarms from the system addressed. The participants were in this case not able to identify the interests that were relevant for the problems, which implies that EthXpert perhaps should, apart from structuring the problem and the relationships between stakeholders, also be enabled to propose possible interests for stakeholders. However, if this was to become the case, the interests should still not be constrained to a set of predefined normative principles or values. This raises an objection against the proposed method. Can a user imagine what the interests of each involved stakeholder are and moreover how these interests affect and are affected by a certain decision? This would be trivial if there existed a true and fixed set of interests that was shared by all rational beings, but, even without claiming moral relativism, I think most of us would agree that it is hard to know exactly what other people value.

It is unsatisfactory to notice that care has been taken to address biases in the process of analyzing how interests are affected by decisions while less care is addressed to supporting the user in selecting a realistic set of interests for the involved stakeholders. Here we definitely risk the fallacy of fundamental attribution error, since it is very human to believe that people in different positions have different interests, while it really would be more reasonable to assume that the polarization of certain inclinations is a situational interpretation of the stakeholder. However, if we do not want to retreat to forming a set of shared values, it does not seem likely that we can find a perfect way to avoid this problem other than to warn users from drawing conclusions without asking for other people's opinions.

On Interests

It has not been explicitly explained above why EthXpert is designed with such a fixation on stakeholder interests. Earlier versions of the P&P method have been less exclusive in the terminology. In those the expressions *principles*, *values*, *duties*, *feelings* and *needs* were used as complementing denominations. There are however both theoretical and practical reasons for using only the simple expression *interest*. The latter is obvious, since a single word reduces the cognitive load of reading and thinking about what different words really imply. The former is related to it. Words like *principles*, *duties* and *values* have a normative connotation and thus assign a sacred, heteronomous, value to the statements. Duties are done unreflectively. Principles and values lie beyond influence. This constrains the process of

scrutiny. Similarly words like *feelings* and *needs* evoke emotions, which puts another type of blanket over the head. All of these expressions can be better operationalized in the word *interests*. This operationalization was done also by Kant, who further distinguished between interests of reason and interests of inclinations, the latter of which would result in good only by accident. It is therefor important that the scrutinizing of interests leads to a questioning of their legitimacy.

On Stakeholder Participation

If collecting as much information as possible about a problem helps to better handle it, why not let everybody who is interested contribute to the process? The idea as such is not bad. Living labs, open source and open innovation projects show that there is much to gain from allowing the public to influence also complex decision processes. For the decision maker, just getting an impression of the amounts of information created by multiple sources would most likely help to create a wider understanding of the problem at hand. However, the difficulty with such an approach is how to digest huge amounts of uncoded data; the problem will turn into a cognitive or computational one. Tools for automatically narrowing down information would have to be based on the software designer's interpretation of moral problems. This poses an apparent risk for confirmation bias as the information is heterogeneous and allows for interpretation. Also, it should not be forgotten that ethical problems are of the type that makes it unsuitable or even counterproductive to ask for everyone's opinion. Some decisions inevitably become uncomfortable, or even unacceptable, for a lot of people in order to protect other values. Not only do people put great value in their own biased inclinations, people sometimes hold opposing values. Putting a decision up for public debate and enabling people to affect the decision could render extreme polarization of opinions (Baron and Spranca, 1997; Brown, 2000). It is thus not an easy choice whether or not to allow the public to be involved in decision making. There is also an important point in letting the decision maker define and analyze interests for other stakeholders. Her feeling of responsibility for making correct judgments is probably increased. This will also reveal any prejudice and biases that affect the analysis. The decision maker would make a decision also without involving an ethical analysis, so by using the tool she is putting her take on other stakeholders' interests up for inspection, which can help bring systematic misjudgment out in the open at an early stage.

Options

In neither of the studies did the participants manage to increase the number of alternative options to consider compared to P&P. This suggests that the tool is not giving any extra benefit in this aspect. Indeed more effort has been put on helping the users to expand the number of stakeholders and interests so it does not come as a complete surprise. Neither is it as critical for the analysis as the identification of stakeholders. Given a real-life scenario, alternative courses of action are more of a technical nature and constitute a better-defined problem once the values involved in the problem have been identified. It thus seems like EthXpert is a tool that can stimulate the adoption of a more autonomous way of handling moral problems but more effort need to be put on helping users to involve and evaluate optional strategies.

Is this moral?

The final objection concerns the non-normative approach in EthXpert. If the tool does not promote good moral values, how can it ever help decision makers make moral decisions? And is there not a great risk that evil decision makers will use the tool to justify morally bad decisions?

This thesis has argued that moral investigation is something that does not need to be guided by moral preaching in order to give a morally defensible outcome. In fact, it suggests that moral investigation is done better without it. Perhaps preaching can result in more conservative decisions but decisions based on as a systematic scrutiny of morally relevant aspects will produce more natural and still defensible solutions. The documents that are produced can not be used to vindicate an immoral decision, since what they reveal is the subjective interpretation of morality of the decision maker. This means that the documents can only be used to vindicate a morally defensible decision, as for each considered decision alternative, the risks and possibilities for the interests of each considered stakeholder is stated. If these statements are inaccurate, incomplete or biased it should be considered as a severe credibility loss for the decision maker.

Is it not so, that in many moral problem situations we, as individuals, do not know instantly what is morally right to do? In such situations we need to explore the facts to find it out. Sometimes someone has deliberated over the same problem before us and can give us some guidance, but other times we have to find the answer all by ourselves, which in some cases means that we might even have to invent moral codes. This is something that we prefer to let philosophers do, but if we do not have access to them, we have a better chance to reach something sustainable if we deploy methods and tools that guide us through a philosophical process. Thus, let us be philosophers!

Future work

There are many aspects of the proposed tool that need to be investigated further. Below I have gathered a few of these that I consider to be important and interesting to approach.

- Evaluate whether the standard set of stakeholders, as well as other methods deployed in Stakeholder Theory can help users of EthXpert to find relevant stakeholders.
- Evaluate the effect from guiding the thoughts of the user by using a more formative language, e.g. by supplying sentence introductions like “Stakeholder X will...” and “Stakeholder X will not...” when she is determining the consequences from different options.
- Explore how to best visualize the connection between similar data and how to use coloring to achieve better overview.
- Explore how to enable the user to associate stakeholders in considerations and how to subsequently visualize this in a comprehensive way. Understanding that stakeholders are occupying the values of other stakeholders is an important part of understanding the dynamics in a real life situation, but if the decision maker systematically would associate stakeholders with other stakeholders, the resulting complexity would be difficult to keep track of. How this should be visualized is yet an open question.
- The tool is prepared to allow data collection of user behavior. It would be interesting to know how the tool is being used in the analysis process. In which order do users add stakeholders and interests? How many of these do they remove before the analysis is concluded? Such questions are possible to answer and would give an indication of how well the tool supports the forming of problem understanding.

Acknowledgements

This research would not have been possible without generous funding from Handelsbanken Research Fundings: Jan Wallanders & Tom Hedelius Stiftelse and Tore Browaldhs Stiftelse. Thank you.

Next, I would like to acknowledge my supervisor, Professor Iordanis Kavathatzopoulos, for patiently introducing me to the psychological theory of moral decision making, the foundation that the thesis rests on. Nothing of this would have been written, was it not for his relentless guidance. I am grateful for his enthusiastic engagement in my many attempts to question the theoretical basis and for his liberal appreciation for dissent. I am looking forward to further discussions on the topic.

My assistant supervisor, Erik Borälv, also deserves much gratitude for his insightful comments and references to relevant topics. Rebecka Janols, Simon Tschirner and Bengt Sandblad read and gave useful feedback that improved the text. Thank you all!

Finally: Takako, thank you for taking care of me. IOU :-*

Bibliography

- Anderson, J. R. (2010). *Cognitive Psychology and Its Implications*, Worth Publishers, New York.
- Baron J., Gowda R., Kunreuthe H. (1993). Attitudes toward managing hazardous waste: What should be cleaned up and who should pay for it? *Risk Analysis*, 13(2), 183-192.
- Baron, J. and Spranca M. (1997). Protected values. *Organizational Behavior & Human Decision Processes*, 70, 1-16.
- Bederson, B. B., Clamage, A. D., Czerwinski, M. P. and Robertson, G. R. (2004). DateLens: A Fisheye Calendar Interface for PDAs. *ACM Transactions on Computer-Human Interaction*, 11(1), 90-119.
- Block, W.; Kinsella, N. S.; Hoppe H (2000). The Second Paradox of Blackmail. *Business Ethics Quarterly*, 10(3), 593-622.
- Boulding, K. E. (1966). The ethics of rational decision. *Management Science*, 12(6), B161-B169.
- Brans, J. P. (2002). Ethics and decision. *European Journal of Operational Research*, 136, 340-352.
- Brown, R. (2000). *Group processes*, Blackwell Publishers Ltd, Malden, Massachusetts.
- Collins, W. R. and Miller, K. W. (1992). Paramedic ethics for computer professionals. *Journal of Systems Software*, 17, 23-38.
- Davidrajuh, R. (2008). A computing system to assist business leaders in making ethical decisions. In M. Oya, R. Uda and C. Yasunobu (Eds.), *Towards sustainable society on ubiquitous networks*, 303-314. Springer: Boston.
- De Martino, B., Kumaran, D., Seymour, B. J. and Dolan, R. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, 313, 684-687.
- Donaldson, T. and Preston L. E (1995). The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. *The Academy of Management Review*, 20(1), 65-91.
- Eriksson S., Helgesson G. and Höglund A. T. (2007). Being, doing, and knowing: Developing ethical competence in health care. *Journal of Academic Ethics*, 5, 207-216.
- Erlandsson, M. and Kavathatzopoulos, I. (2005). Autonomy method: Acquiring skills for ethical analysis of computerisation in car driving. In G. Collste, S. O. Hansson, S. Rogerson and T. W. Bynum (Eds.), *ETHICOMP 2005: Looking back to the future* [CD-ROM], Linköping University, Linköping.
- Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology*, 1(1), 33-52.
- Foucault M. (1987). Panoptismen. In *Övervakning och straff – Fängelsets födelse*, 196-228. Arkiv, Lund.
- Friedman, B., Kahn, P. H., and Borning, A. (2008). Value Sensitive Design and information systems. In K. E. Himma and H. T. Tavani (Eds.), *The Handbook of*

- Information and Computer Ethics*, 69-101. John Wiley & Sons, Inc., Hoboken, NJ.
- Galinsky, A. D., Maddux, W., Gilin, D. and White, J. B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science*, 19, 78-84.
- Google (2010). Available through: Google <<http://www.google.com/search?q=%22stakeholder+analysis+tool%22>> [Accessed 2010-08-18].
- Gotterbarn, D. W. (2002). Reducing software failures: Addressing the ethical risks of the software development lifecycle. *Australian Journal of Information Systems*, 9(2), 155-165.
- Greene, J. and Haidt, J. (2002). How (and where) does moral judgment work? *TRENDS in Cognitive Sciences*, 6(12), 517-523.
- Gustafsson, C. (1997). Den lystna maskinen [The greedy machine]. In A. Björnsson and P. Luthersson (Eds.), *Medialiseringen av Sverige*, 63-75. Carlsson, Stockholm.
- Kahneman, D. (2002/2010). Maps of bounded rationality: a perspective on intuitive judgment and choice, Prize Lecture [pdf]. Available through: [Nobelprize.org](http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-lecture.html). <http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-lecture.html> [Accessed 2010-08-26].
- Jetter, H-C., Gerken, J., König, W., Grün C. and Reiterer H. (2006). HyperGrid - Accessing Complex Information Spaces. *People and Computers XIX - The Bigger Picture 2006*, Part 3, 349-364.
- Kant, I. (1785/2010). *Fundamental Principles of the Metaphysic of Morals* [e-text]. Available through: Project Gutenberg <<http://www.gutenberg.org/etext/5682>> [Accessed 2010-08-26].
- Kavathatzopoulos, I. (1993). Development of a cognitive skill in solving business ethics problems: The effect of instruction. *Journal of Business Ethics*, 12, 379-386.
- Kavathatzopoulos, I. (1994). Training professional managers in decision-making about real life business ethics problems: The acquisition of the autonomous problem-solving skill. *Journal of Business Ethics*, 13, 379-386.
- Kavathatzopoulos, I. (2003). The use of information technology in the training for ethical competence in business. *Journal of Business Ethics*, 48, 43-51.
- Kavathatzopoulos, I. (2004). Making ethical decisions in professional life. In H. Montgomery, R. Lipshitz and B. Brehmer (Eds.), *How professionals make decisions*, 277-288. Lawrence Erlbaum Associates, Inc, Mahwah, NJ.
- Kavathatzopoulos, I., Laaksoharju, M. and Rick, C. (2007). Simulation and support in ethical decision making. In T. W. Bynum, K. Murata and R. Simon (Eds.) *Globalisation: Bridging the global nature of Information and Communication Technology and the local nature of human beings*, 278-287. Meiji University, Tokyo.
- Kavathatzopoulos, I. and Rigas, G. (1998). A Piagetian scale for the measurement of ethical competence in politics. *Educational and Psychological Measurement*, 58, 791-803.
- Kavathatzopoulos, I. and Rigas, G. (2006). A measurement model for ethical competence in business. *Journal of Business Ethics Education*, 3, 55-74.
- Klein, P, Müller, F, Reiterer, H. and Eibl, M. (2002). Visual Information Retrieval with the SuperTable + Scatterplot. In E. Banissi (Ed.), *Proceedings of the Sixth International Conference on Information Visualization (IV'02)*, IEEE Computer Society Press, 70-75.

- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. and Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, 446, 865–866.
- Kohlberg, L. (1984). *Essays on moral development, Vol. II: The psychology of moral development*, San Francisco, CA: Harper & Row.
- Kohlberg, L. (1985). The Just Community: Approach to moral education in theory and practice. In M. Berkowitz and F. Oser (Eds.), *Moral Education: Theory and Application*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Kunsch, P. L., Kavathatzopoulos, I. and Rauschmayer, F. (2009). Modelling complex ethical decision problems with Operations Research. *Omega - The International Journal of Management Science*, 1100-1108.
- Laaksoharju, M. (2008). *Ethicking - Dealing with Ethics. Can Micro World Simulations Stimulate Ethical Competence?* UPTec IT08 002, Uppsala Universitet, Uppsala.
- Laaksoharju, M. and Kavathatzopoulos, I. (2008). Can micro world simulations assess and stimulate ethical competence? in T. Ward Bynum, M. Calzarossa, I. De Lotto and S. Rogerson (Eds.), *Living, Working and Learning Beyond Technology*. Mantova, Italy: Tipografia Commerciale (pp. 503-510).
- Mancherjee, K. and Sodan, A. (2004). Can computer tools support ethical decision making? *ACM SIGCAS Computers and Society* [online], 34(2). Available through: ACM <<http://portal.acm.org/citation.cfm?id=1052792>> [Accessed 2010-08-30].
- Maner, W. (2002). Heuristic methods for computer ethics. *Metaphilosophy*, 33, 339-365.
- Moodley, K., Smith, N. and Preece C. N. (2008). Stakeholder matrix for ethical relationships in the construction industry, *Construction Management and Economics* 26, 625–632.
- Piaget, J. (1932). *The moral judgement of the child*. Routledge and Kegan Paul, London.
- Ragatz, L.L and Russell, B. (2010). Sex, Sexual Orientation, and Sexism: What Influence Do These Factors Have on Verdicts in a Crime-of-Passion Case? In *The Journal of Social Psychology*, 150(4), 341–360.
- Rao, R. and Card, S. K. (1994). The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus and Context Visualization for Tabular Information. In B. Adelson, S. Dumais and J. Olson (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence (CHI'94)*, 318-322. ACM Press.
- Rauschmayer, F., Kavathatzopoulos, I., Kunsch, P. L. and Le Menestrel, M. (2009). Why good practice of OR is not enough – ethical challenges for the OR practitioner. *Omega - The International Journal of Management Science*, 37, 1089-1099.
- Jacko, J.A, Sears, A. (2003). *Human-Computer Interaction Handbook*, Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.
- Schultz, M. (2010). Ett försvar för incest [online]. Available through: *Mårten Schultz – Sveriges största juridikblogg* <<http://martenschultz.wordpress.com/2010/03/16/ett-forsvar-for-incest/>> [Accessed 2010-08-07]
- Sloman, S. A. (2002). Two systems of reasoning, in T. Gilovich, D. Griffin and D. Kahneman (Eds.), *Heuristics and Biases: the Psychology of Intuitive Judgment*, 379-396. Cambridge University Press, Cambridge.

- Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences*, 28, 531-573.
- Wenstøp, F. and A. Myrmel (2006). Structuring organizational value statements. *Management Research News*, 29(11), 673-683.

Recent licentiate theses from the Department of Information Technology

- 2010-004** Kenneth Duru: *Perfectly Matched Layers for Second Order Wave Equations*
- 2010-003** Salman Zubair Toor: *Managing Applications and Data in Distributed Computing Infrastructures*
- 2010-002** Carl Nettelblad: *Using Markov Models and a Stochastic Lipschitz Condition for Genetic Analyses*
- 2010-001** Anna Nissen: *Absorbing Boundary Techniques for the Time-dependent Schrödinger Equation*
- 2009-005** Martin Kronbichler: *Numerical Methods for the Navier-Stokes Equations Applied to Turbulent Flow and to Multi-Phase Flow*
- 2009-004** Katharina Kormann: *Numerical Methods for Quantum Molecular Dynamics*
- 2009-003** Marta Lárusdóttir: *Listen to Your Users - The Effect of Usability Evaluation on Software Development Practice*
- 2009-002** Elina Eriksson: *Making Sense of Usability - Organizational Change and Sense-making when Introducing User-Centred Systems Design in Public Authorities*
- 2009-001** Joakim Eriksson: *Detailed Simulation of Heterogeneous Wireless Sensor Networks*
- 2008-003** Andreas Hellander: *Numerical Simulation of Well Stirred Biochemical Reaction Networks Governed by the Master Equation*
- 2008-002** Ioana Rodhe: *Query Authentication and Data Confidentiality in Wireless Sensor Networks*
- 2008-001** Mattias Wiggberg: *Unwinding Processes in Computer Science Student Projects*
- 2007-006** Björn Halvarsson: *Interaction Analysis and Control of Bioreactors for Nitrogen Removal*
- 2007-005** Mahen Jayawardena: *Parallel Algorithms and Implementations for Genetic Analysis of Quantitative Traits*
- 2007-004** Olof Rensfelt: *Tools and Methods for Evaluation of Overlay Networks*



UPPSALA
UNIVERSITET

Department of Information Technology, Uppsala University, Sweden