Theses in Philosophy from the Royal Institute of Technology 24

Trusting and Taking Risks

A Philosophical Inquiry

Madeleine Hayenhjelm

STOCKHOLM 2007

Division of Philosophy
Department of Philosophy and the History of Technology.
Royal Institute of Technology
SE-100 44 Stockholm, Sweden.
Abstract


This dissertation is a philosophical contribution to the theories on trust and on risk communication. The importance of trust in risk communication has been argued for and empirically studied since the 80s. However, there is little agreement on the notion of trust and the precise function of trust. This thesis sets out to study both aspects from a philosophical point of view. The dissertation consists of five essays and an introduction.

Essay I is a comment on risk perception theory and the psychometric model in particular. The essay argues that sometimes individuals take risks simply because they are in a vulnerable position and have no other choice. Four factors are identified as crucial in risks from vulnerability: poor outset conditions, lack of reasonable options, hope and liability to disinformation.

Essay II addresses the democratic approach to risk communication and the ideal of risk communication as a dialogue between more or less equal stakeholders. The essay develops a typology of different risk communication practices and influence distributions and concludes that the form of risk communication most commonly referred to in the literature is of the most imbalanced kind where all three types of influence falls to the same party.

Essay III argues that trust can be understood as *comfortable defocusing*. Trust as comfortable defocusing means that when a person trusts, he or she is comfortable with letting certain things be in the hands or in the control of the trusted other. Departing from this idea a definition of trust is proposed consisting of three elements: (i) the truster’s comfortable defocusing from the trusted person’s part of caring for X, based on a (ii) belief about the trusted person’s care for X, concerning (iii) something of concern to the truster.

Essay IV explores the concept of trustworthiness and under which conditions trust can be betrayed in a morally justified way. It is argued that two aspects are essential for every reasonable notion of trustworthiness: reliability in terms of commitment to trust and predictability in terms of compliance with norms. In order to be untrustworthy in a morally justified way two major conditions need to be fulfilled: first, what is expected from the trustee needs to be either immoral or part of an ongoing immoral or unjust relationship between truster and trustee and, secondly, the trustee must not be in a position to explicitly reject trust.

Essay V examines two of the most employed distinctions within the trust literature, that of trust/confidence and of trust/reliance. These distinctions are evaluated against a set of five criteria. It is argued that both these distinctions are employed to cover several distinguishing properties, thus adding to conceptual confusion.

**Key words:** beliefs, comfortable defocusing, disposition, power, public trust, risk, risk communication, trust, trustworthiness, untrustworthiness, vulnerability.
List of Essays

This doctoral thesis consists of the following essays preceded by an introduction:


Hayenhjelm, Madeleine (2007) – “Un-confusing the Notion of Trust”, submitted manuscript.

Essay I is reprinted with kind permission of *Journal of Risk Research* for Taylor and Francis Publishers. See www.taylorandfrancis.co.uk.

Acknowledgements

This dissertation would not have been what it is without the help and encouragement of a range of friendly philosophical souls. My first and deepest thank goes to my supervisor Professor Sven Ove Hansson and my assistant supervisor Martin Peterson. I would also like to thank Lennart Sjöberg, Misse Wester Herber, Nils-Eric Sahlin, Marion Ledwig, John Cantwell and all my colleagues at the Ethics and Philosophy of Risk seminars at the Philosophy Department at the Royal Institute for Technology for valuable input and comments. Lars Lindblom, Birgitte Wandall, Mikael Dubois, Sara Svensson, Kalle Grill, and my roommate Hélène Hermansson, all deserve a special mention for contributing to this thesis in various ways.

Stockholm
August 2007
Madeleine Hayenhjelm
# Table of contents

**ABSTRACT**

**ACKNOWLEDGEMENTS**

**INTRODUCTION**
- Introduction
  - What is Risk Communication?
  - The Significance of Trust to Risk Communication
  - The Notion of Trust in Risk Research
  - The Notion of Trust in Moral Philosophy
  - The Combined Perspective

**Preview of Essays**

**References**

**ESSAY I:**
- Out of the Ashes: Hope and Vulnerability as Explanatory Factors in Individual Risk Taking

**ESSAY II:**
- Asymmetries in Risk Communication,

**ESSAY III:**
- Trust as Comfortable Defocusing
  - Submitted manuscript

**ESSAY IV:**
- Morally Justified Untrustworthiness
  - Submitted manuscript.

**ESSAY V:**
- Un-Confusing the Notion of Trust
  - Submitted manuscript.
Introduction

This is a philosophical thesis about trust and the communication of risks. Trust has been widely addressed and explored within contemporary moral philosophy, particularly since Annette Baier wrote her seminal paper on trust in the 80s (Baier 1986). The notion of trust has also been applied, as in Onora O’Neill’s work on trust and bioethics (O’Neill 2002a). During the same period, from the 80s and throughout the 90s, trust has received considerable attention within risk research, in empirical research on public risk perception and in communication strategies, and also in the theoretical framework informing such research and practices. The risk concept of trust draws primarily from sociological and psychological resources, such as Luhmann’s (1979) and Giddens (1990), as well as from empirical findings from psychometric studies. The need for a clarified concept of trust becomes particularly accentuated in empirical studies of trust where a consistent and clear notion is a prerequisite for conclusions about its impact and causes. The philosophical literature on trust, although clarifying on a number of points, is still rather divided on what kind of phenomenon trust is. Debatable issues include to what extent trust can be chosen, to what extent it can be unconscious, and to what extent it can include a precise idea of what the trusted person is expected to do. It is from these two distinct framework and perspectives that this thesis departs. Its aim is to contribute to both literatures.

The thesis addresses two over-arching topics. The first is to clarify the notion of trust in such a way that it both contributes to the conceptual literature on trust in philosophy and is applicable to risk research. The second is to clarify the role of public trust in the risk communication context. Trust has been given a number of roles within risk research as a prerequisite for communication, as a moderator, and as an objective for communication. The intended object for trust varies, and can be sources for information, decision makers and the message itself. In order to explain the role trust is expected to take in risk communication, the multiple forms of risk communication and the implied relations between experts and the public need further investigation.
Risk communication mainly refers to dissemination of science-based information about risk and policy-based decisions about risk, to a lay audience. Such information could be distributed in the form of public campaigns with the aim of reducing public risk behaviour, such as smoking. It could also be in the form of a dialogue between the local public and, for example, representatives from a chemical factory discussing the waste programme of the plant. Risk communication can also refer to more or less any kind of communication between experts, among laypersons, etc, concerning risk. However, imbedded in the idea of risk communication there is a much more general and abstract idea about the possession, and dissemination, of information about possible negative outcomes. Such an idea, in this abstract sense, seems to point beyond the applied risk field to moral philosophy and more general questions concerning moral obligations with regard to information in general and with regard to information about possible negative consequences for actions in particular.

There are at least two things that there seem to be almost a consensus about within risk research concerning trust. First, that trust is important for risk perception, risk acceptance and risk communication (see e.g. Siegrist et al. 2003, 705, 707). Second, that there is no agreement on how trust ought to be conceptualised, or what its precise function is, and that the very multitude of proposed theories and concepts make the whole field rather confused (Cvetkovich and Winter 2007, 188; Siegrist et al. 2003, 705f; Earle et al. 2007, 3; Löfstedt 2005, 6). Cvetkovich and Nakayachi comment on trust research within the social sciences, including risk management, thus: ”The field has been characterized as a theoretical quagmire and a verbal and conceptual morass marked by a confusing potpourri of definitions…” (2007, 224).

There are basically two ways to address such a field lacking in previous philosophical discussion. One can approach the subject from the general philosophical literature and apply well-known philosophers views on vaguely related matters to the new field. Or, one can look at the theories of field in question and try to disentangle the conceptual and theoretical knots, propose new concepts, and modify the theories. The first approach calls for a bit of caution however, since particularly Habermas (see e.g. Rowan 1994, 401 and Kemp 120f. For critical comment see e.g. Gabrill and Simmons 1998, 423), but also Aristotle (Peters et al, 1997, 47) Nussbaum (Earle and Cvetkovich
1995, esp. Ch. 10) and others (see e.g. Earle and Cvetkovich 1994) have all been employed to support one heuristic framework over another in the risk literature, often with little more than a vague reference to a whole work. This thesis draws from existing philosophical theory about trust as well as the risk related literature on trust, treating both literatures as theoretical proposals worth scrutinizing philosophically. I have approached the matter, as I would have any field of inquiry: asked what the relevant questions are; looked for essential aspects which may have been overlooked; asked whether the existing concepts and theory cover what they claim to cover, etc. I have thus addressed the risk communication literature as if it was one of philosophy and applied philosophical criteria such as conceptual consistency, theoretical exhaustiveness, separating the theoretically necessary from the contingent, etc.

This introduction is divided into seven parts. The next section will shed some light on how risk communication is conceptualised in the risk literature. The subsequent section describes common ideas about the role of trust in risk research. The forth and fifth section give an outline of common ideas of the notion of trust in the risk literature and the philosophical literature, respectively. The sixth section is a concluding discussion and also a background to the contributing papers. The eighth part consists in short previews of the papers included in the thesis.

The dissertation consists of five papers. The first two papers are contextual: they primarily scrutinize the received assumptions concerning the empirical conditions and constraints for trust with regard to risk decisions. The first paper situates individual risk taking in a context of available options. The second paper situates risk communication in a situation of inequality between decisions makers and those affected by the risks. The implied thought here is that the risk taking involved in trusting someone can be understood in a context of potential vulnerability, either due to lack of choices or lack of relative influence. The subsequent three papers are conceptual explorations of the concept of trust, and its sister concept trustworthiness, as such. The third paper develops a theory about trust and attention as a contribution to the existing literature on trust as belief and attitude. The fourth paper completes the above account of trust with an account of trustworthiness based on norms, and explains how trustworthiness can both be of moral value and at times be a morally good thing to not to live up to. The
fifth paper brings the two perspectives, philosophy and risk management, together in a conceptual discussion of two popular distinctions in each discipline, and attempts at clarifying the trust notion as conceived in the risk literature.

**What is Risk Communication?**

In order to address the question of the importance of trust to risk communication we need to sort out what risk communication is. The term ‘risk communication’ was first formulated in 1984 (Gurabardhi *et al* 2005). Risk communication research evolved in the 1980s (Chess *et al* 1995) as a development from the risk perception research that had developed in the 70s and in response to public protests against controversial technologies such as nuclear power (Grabill and Simmons 1998). This means that risk communication has right from the start developed as a means for addressing the discrepancies between lay and expert views on risks. Since then two different approaches of how to achieve a more common understanding between experts and public have evolved (see e.g. Rowan 1994). The first is the top-down or technocratic approach that addresses risk communication as experts informing the public about risks in order to influence and persuade the public to share the expert’s view. The other approach is the participatory or stakeholder involvement approach that stresses a two-way communication or a mutual exchange of information between experts and public. Behind this approach lies an assumption that risk decisions are not only technical decisions but are inevitably also rather value-laden. Since the public is as much experts on values as the scientific experts, risk communication should aim for mutual learning and understanding. In its widest sense, as hinted at in the introduction, risk communication can be conceived of as information about negative outcomes with regard to difficult decisions and the distribution or exchange of such information. Such a conception is however too wide to grasp the implied meaning in this dissertation and the literature it draws from. There are several ways to explain risk communication that are part of the self-conceptualisation within the risk management field.

One approach is to simply list all activities referred to as ‘risk communication’. Such a list generally includes warning labels, doctor-patient conversations, public health
campaigns, citizen advisory panels, surveys, etc. In other words, it includes all examples of professionally distributed oral or written material about risks to health, safety or environment, or conducted in the forms of forums or dialogues with the public. Typically the recipient is a member of the public and typically the source is scientific experts, regulators or industry. Such lists are often found in a theoretical context; often either risk related, or management related.

Another way to describe risk communication is to situate it within the context of risk analysis and the management of risk. In such descriptions risk communication is the dissemination of information about the result of risk assessments and risk management decisions. This view of risk communication as dissemination of pre-decided risk decisions and expert views has been criticised, and more participatory approaches to communication have become increasingly influential. There is a developmental narrative often referred to from the top-down communications of numbers in the past, to more participatory processes involving the public as partners, aiming at shared values and mutual trust in more recent communication. Regardless of whether risk communication is a two-way dialogue aiming at joint decision-making or a large organization informing the public about its risk decisions, the context is one of large risk decisions and communication between the general public and the organization in charge.

Risk communication can also be linked more specifically to the psychological studies of public risk perception (see Fischhoff et al 1978, Slovic 1987). It was found that experts and the general public perceived risks differently, the public taking a wider set of concerns into consideration than the experts (see e.g. Slovic et al 1980; Slovic 1986, Slovic 1987). The public would also take aspects such as degree of control, catastrophic potential and familiarity, etc into their accounts of the dangers of various sources of risks, whereas experts would merely look at probabilities for death or other negative outcomes (see e.g. Fischhoff et al 1978). Common to the risk research field is the idea of a "knowledge gap", or an "information deficit", that risk communication was supposed to fill (Frewer & Salter 2007, 145, Pigeon et al 2007, 118; see also Leiss 2004). It was thought that communication would close the gap between the scientific understanding of risk and that of the public. This understanding of risk communication partly coincides with the one above. However, is not restricted to industrial or technical risks but also
includes information from regulators and government agencies concerning individual risk taking.

Another approach is to describe risk communication as a professional field of communication with the public and thus part of public relations and management theory, rather than risk analysis. Some of the activities referred to as risk communication may just as well be described in terms of public relations, stakeholder theory, public outreach programmes, etc. Risk communication is then conceptualised and understood as part of the wider field of corporate management of relations with the public. Besides the links to organization theory, management theory, and communication theory, there are communication practitioners who define risk communication as a method for effectively communicating with the public about risks based upon practical guidelines and heuristics developed from the practices in the field. According to Covello et al (2001) “Effective risk communication is a professional discipline; its application requires knowledge, planning, preparation, skill, and practice.” They go on to describe effective risk communication as a two-way process where the lay public learns about a particular risk, and the risk assessment and management of that risk, and the risk experts learn about the interests and concern of the public.¹

Finally, risk communication can be described by referring to one or more of the theoretically informed definitions. These typically try to fit all of the referred practices into a common concept that at the very least comprises both participatory and disseminating information practices, together with public health campaigns, and emergency information. One such example is the much quoted definition of risk communication by Covello, von Winterfeldt, and Slovic (1986):²

…any purposeful exchange of information about health and environment risks between interested parties. More specifically, risk communication is the act of conveying or transmitting information between parties about: a) levels of health or environmental risks; b) the significance or meaning of health or

¹ See also Leiss (2004, 401) for a similar view of risk communication: “Risk communication is the process of communicating responsibly and effectively about the risk factors associated with industrial technologies, natural hazards, and human activities.”

environmental risks; or c) decisions, actions or policies aimed at managing or controlling health or environmental risks (Covello et al 1986, 172)

Other theoretically inclined explanations list the different aims of risk communication. Typically such a list includes the aim to affect public risk behaviour; to inform and educate the public; and to establish an improved relation with between the public and risk managers, in terms of joint decision making, cooperation or mutual trust or in terms of acceptance of risk decisions. As an example, the list of different functions of risk communication proposed by the OECD can be mentioned (cited in Hampel 2006, 8):

1. Education and enlightenment
2. Risk training and inducement of behavioural changes
3. Confidence in institutions of risk assessment and risk management
4. Involvement in risk related decisions and conflict resolution

Covello *et al* (2001, 383) mention three similar objectives: “providing the knowledge needed for informed decision making about risks; building or rebuilding trust among stakeholders; and engaging stakeholders in dialogue aimed at resolving disputes and reaching consensus.”

In this thesis the focus is on risk communication as a means to establish a supportive relation between the public and risk decision makers, and to make information concerning risks in individual risk taking available.

**The Importance of Trust to Risk Management**

The interest in the notion of trust within risk communication theory can be traced to research in risk perception in the 1980s, concerning the so called ‘knowledge gap’ between risk perception of experts versus that of the general public (see Poortinga and Pidgeon 2003, 961). It was suggested that this gap could be explained in terms of the perceived trustworthiness of risk managers. Later risk research studies found that low trust in regulators was correlated to perceived high risk with regards to the risk sources
that the regulators controlled, and Slovic (1997) claimed in a seminal paper that trust was more important than risk communication for influencing risk perception. Besides the suggested correlation between risk perception and trust, this finding has been taken to have implications for risk communication and risk acceptance.

In general, trust in risk management is negatively related to risk perception. This is an important observation because it opens a possible pathway to affecting public risk perception and improving risk management: if we understood trust, and if we could affect levels of trust, then we might also be able to affect levels of risk perception and, ultimately, risk acceptance/rejection. (Earle et al. 2007, 1)

There is thus an idea of a causal link from trust in risk managers to low risk perception to risk acceptance. In other words if risk communication can achieve public trust in risk managers or risk regulators, then this will affect how dangerous a certain risk will be perceived. If a certain risk is perceived as less dangerous and this accords with the judgement of risk managers, this will lead to greater acceptance to the decisions that risk managers make concerning that risk.

Parallel to the findings in risk perception research on trust, there has also been a concern for indications of a decline in public trust in government agencies and industry (see e.g. Peters et al 1997, 43; Leiss 2004, 392; Löfstedt 2005, 9) and the awareness of distrust being a problem for communicating risks (see Petts 1998, 307; Löfstedt 2003, 419). Löfstedt (2003, 418), for example, claims that the various risk communication programmes have largely been ineffective, and suggests among other things, that increased attention to trust and public participation is needed.

This is where the whole topic becomes complicated. The rise in public distrust is seen as a problem that can be remedied by risk communication. Risk communication is then the means and trust is the goal. But there is also the other idea that risk communication fails because there is too little trust, thus trust needs to be established first, so that other objectives of risk communication can later be achieved (see e.g. Covello et al 2001, 386). Poortinga’s and Pidgeon’s view may be regarded as typical:
Also, in the field of risk research there is now general agreement that trust in risk management institutions may be an important factor in risk perception and acceptance of risks [...] Furthermore, trust is seen as one prerequisite for effective risk communication, while distrust may be associated with stigmatization of technologies, such as nuclear power as well as social amplification effects following major failures of risk regulation. (Poortinga and Pidgeon, 2003, 961)

Alongside the new interest in trust in the risk literature from the late 80s and early 90s, there was also an increased interest in dialogue forms for risk communication, since the public perception of risk remained unaffected by top-down communication (see e.g. Löfstedt 2005, 19). These more participatory forms of risk communication were seen as a way to increase public trust and learn more about public interests and values.

Later research has also indicated that besides influencing how a delivered risk message is received, trust may also affect which sources a person chooses to turn to for information. Alaszewski (2005, 103) stresses trust as important for which sources of information are influential to risk perception. Petts (1998) point to trust as influencing how information is sought, as well as received:

Trust has important implication at the interface between the expert/decision maker and the public, for example, impacting upon: (i) who people look to for information, (ii) the response of the public to technical information and its providers, (iii) the manner in which consultation and communication is conducted, (iv) the extent to which control of a risky activity is thought to be achievable. (Petts 1998, 307).

We can then distinguish at least three different but interrelated roles for public trust with regard to risk management. First, there is the idea that trust in risk managers affects risk perception. This idea has been widely discussed and studied in risk perception literature along with other factors that are taken to influence how risks are perceived. Secondly, there is the idea that one of the main objectives for risk communication is to establish trust in the risk managers. The goal is to achieve public trust in risk decisions made and
risk communication is the means. In this context there is slight shift in the objective from risk perception to risk acceptance. Thirdly, it is assumed that in order to establish risk communication there is need to first establish trust. These three role for trust can be summarized thus:

1. Trust is an important factor in risk perception, i.e. how the public regard certain risks.
2. Trust in organizations is one of the key objectives for risk communication, in order to achieve cooperation and risk acceptance.
3. Trust is necessary for risk communication to be successful and in order to get information across to the public.

In this thesis I will, to the extent that it is possible to separate the three, focus on trust in risk managers and risk decisions as an objective for risk communication, rather than on trust informing risk perception or trust affecting how information received. I would in particular like to point to the difference between affecting risk perception and affecting acceptance of risk decisions. Affecting the first is closely related to the objective of informing and educating the public from an expert’s point of view. An example of this could be trying to affect risk behaviour with regard to health choices or fear that the scientific community finds unwarranted. Risk acceptance on the other hand is a goal in the interests of risk managers, whether regulators or industry, amounting to public support for risk decisions. The point is that if risk decision makers are perceived as trustworthy, their decisions may be accepted also by a public that does not share the views of the decision makers about the risk in question. Löfstedt writes:

'Trust can be seen as a 'complexity reduction thesis': that is, 'trust' means acceptance of decisions by the constituents without questioning the rationale behind them. In such a case constituents are in effect asking to accept a 'risk judgement' made by the regulators. In using this definition, trust becomes something regulators should strive for. (2005, 6).
It is from this perspective that the notion of trust is explored in the present thesis. This raises a range of normative issues: To what extent is public trust in risk decisions desirable? What is the value of being trusted by the public? etc. But most of all: What is this phenomenon named trust? To what extent does it contribute to cooperation and risk acceptance? What would be the case if it were true of a certain population that it trusted a certain organization in charge of large-scale risk decisions? Would that be a desired state of affairs? What are the alternative scenarios, besides public risk acceptance and public distrust?

The Notion of Trust in Risk Research

Within risk research there are on the one hand a rather large number of empirical studies of trust with a rather un-theoretical notion of trust, and on the other a rather large number of proposals for how trust ought to be conceptualized. The notion of trust within risk research has primarily been employed for empirical quantitative studies. The focus has then been on identifying factors that could increase or decrease trust in a particular source or organization. This means that some studies have identified trust with a number of sentences that the researchers have taken to bring out the meaning of trust, others have left the meaning of trust to the research subjects and asked for them to rate their trust in different sources, and others still have drawn from a particular theoretical framework of trust.

Among the conceptual proposals in the more theoretical risk literature there are some approaches that re-occur in the literature in various forms. One of the most influential approaches to conceptualize trust is to refer to it as multi-dimensional and then list a number of characteristics that are assumed to capture how the truster perceives the trusted other (see e.g. Peters et al 1997, 43-44; Pidgeon et al 2007). This is commonly referred to as the “dimensions” approach. Several studies are occupied with the number of relevant dimensions, whether trust is two-dimensional or more. This approach basically describes trust in terms of expected or desired qualities of the trusted. Peters et al (1997) for example, define trust in terms of knowledge and expertise; openness and honesty; and concern and care. Löfstedt’s (2005) notion of trust is based on perceived fairness, competence and efficiency. Generally there is at least one
dimension referring to competence and skills of some sort, and at least one dimension referring to a normative element such as caring, honesty, fairness, etc. There have been attempts to fit more multi-dimensional theories into bi-dimensional ones, and discussions concerning the appropriate number of dimensions (see e.g. Poortinga and Pidgeon 2003).

A different approach refers to the sociological and psychological literature on trust and defines the notion by characterizing the state of mind of the truster. The definition by Rousseau et al has been rather influential. They define trust in terms of accepting vulnerability: “Trust is a psychological state compromising the intention to accept vulnerability based upon positive expectations of the intentions or behaviour of another” (Rousseau et al 1998, 395). Sociologist Niklas Luhmann’s theories on trust have been equally influential, in particular Luhmann’s distinction between ‘trust’ and ‘confidence’ (Luhmann 1988) and his idea of trust (and confidence) as ‘reduction of social complexity’ (Luhmann 1979) (see e.g. Siegrist et al 2003, 712; Löfstedt 2005, 6).

One prominent theory about trust in risk management is the TCC model (trust confidence cooperation), developed by Earle, Siegrist and Gutscher (Earle et al 2007; Siegrist et al 2003). This model builds on both Rousseau et al and Luhmann and assumes that trust and confidence can contribute to cooperation, i.e. risk acceptance, through two separate routes of rationales and thus two different kinds of information. Trust is then defined as a willingness to make oneself vulnerable to the other based on perceived similarity in values and intentions. Confidence is defined as the belief that certain future events will occur as expected.

The dimensionality approach does however not provide a proper definition of trust, since what is defined is trustworthiness and not trust. It is merely presumed that trust is the same as perceiving someone as trustworthy, that the belief in another person’s trustworthiness is a sufficient and necessary condition for trust. Thus, if a person believes that a risk manager is efficient, fair, and competent, then he or she does per definition trust him or her. Distrust would then be to believe that the risk manager was not competent, or not fair, or not competent. It seems that this confuses trust with what would be a good reason to trust. It is at least conceivable that a person can trust another without believing that he or she is fair, or competent, or efficient. Löfstedt tries
to step around this problem by proposing that trust can either be directed toward the outcome of a decision (to be fair, competent and efficient) or toward the process (to be fair, competent and efficient).

The truster-descriptive account found in Earle, Siegrist, and Gutscher has the value of defining trust rather than trustworthiness (of the trusted person). The explanatory value of this model is however presumed to cover also previous multi- and dual-dimensions model for trustworthiness, so that the model not only distinguishes between two kinds of trust (in the truster) but also between all the different aspects of trustworthiness. ‘Caring’ and ‘honesty’ for example are then matched with trust, whereas ‘competence’ and ‘skills’ are matched with confidence. But why all different aspects of trustworthiness ought to equal exactly two distinct kinds of trust needs to be argued for. This seems to indicate that for every fundamental kind of trustworthiness there is a matching kind of trusting frame of mind. Trust is then reduced to the belief in a particular kind of trustworthiness.

The Notion of Trust in Moral Philosophy

The concept of trust has a parallel discussion within moral philosophy. There are basically two different lines of thought in the contemporary trust research in moral philosophy. The first stems from Baier’s seminal paper ”Trust and Antitrust” (Baier 1986) and her Tanner Lecture on Human Values on trust, (Baier 1992a, 1992b) which pointed to the need to address the topic of trust in moral philosophy. Her account of trust has more or less influenced all later discussion on trust. A different line of thought concerning trust can be found in Hertzberg (1988) and Lagerspetz (1996) in the Wittgensteinian tradition, drawing from Wittgenstein’s On Certainty. A separate discussion of trust, related to testimony, is found in epistemology but has not been addressed in this dissertation (see e.g. Hardwig 1991).

Baier re-introduced the question of trust into the moral discussion, pointing to trust as a kind of acceptance of vulnerability. In contrast to ideas of the contract, trust is often not in the form of a mutual agreement between two consenting parties, but can be implicit, un-chosen or something that had has developed over time. Trust, on Baier’s account, is reliance on the goodwill of another to care for something that the truster
values. The reliance on the motivations of the trusted other is central to her notion of trust. Her distinction between ‘trust’ and ‘reliance’, her view of trust as implying vulnerability, her view of an emotional aspect of trust, as well as her views on rational trust, have all been influential to the literature.

Most of the philosophical literature on trust has been devoted to capturing the phenomenon itself and its defining psychological characteristics. Two primary concerns in the literature have been to identify, and assess the relative weight of, the rational and emotional aspects of trust. On the one hand it is argued that trust is an emotional stance that we take toward another person when trusting; we regard the person we trust in a certain light (See e.g. Becker 1996; Morgan 1973; Holton 1994; Lahno 2001). This attitude toward the other may make us believe certain things about him or her as a result of trust but is not itself a belief (Becker 1996). On the other hand, trust is described as expectations and beliefs about the trustworthiness or the motives of the other that we are optimistic about (see Jones 1996, Baier 1986, Pettit 1996, Mullin 2005). Where trust is placed on this scale has implications for the extent to which trust can be explained in terms of expectations, whether it can be a choice or not; and to what extent it can be based upon reasons.

The role of expectations in trust poses a problem. To trust someone seems on the one hand to be best explained as positive expectations about the future actions or motives of that person. On the other hand, in most situations characterized as trust, the truster does not seem to have any particular thought or belief about the trusted other at all (Lagerspetz 1996). Baier (1986, 235) allows for a notion of trust in certain passages not as positive expectations but as a lack of negative ones. To trust can be to not expect acts from ill will or acts from motives not directed toward the truster. Also Govier describes trust this way: “There are many terrible things people can do to each other, and when we trust a friend we assume confidently – without thinking of it – that she will not do any of these things. We assume that she will not break confidences, tell our secrets, abuse our children, or try to steal our job” (Govier 1997, 4). This makes trust seem like a rather peculiar phenomena being the lack of negative expectations. Not expecting ill will is not a state of mind, as Lagerspetz (1996) has pointed out.
Lagerspetz’s solution is to refer to expectations as a description from a third person perspective that is revealed in behaviour rather than in terms of beliefs and feelings.

It seems that any account of trust ought to be able to explain the kind of everyday trust that does not take the form of calculated or articulated thoughts about what the other will or will not do. This does not exclude a more cognitive kind of trust where a person explicitly entrusts something to another to care for, as when lending or delegating something to somebody. But such trust does not cover all or perhaps not even most cases. Thus any account of trust that simply lists beliefs about what the other is expected, or not expected to do neither exhausts the meaning of trust, nor captures a necessary property.

The Combined Perspective

There is a tendency common to both the philosophical literature and risk literature to not distinguish between trust and trustworthiness. Definitions of trust are often proposed in terms of positive expectations or attitudes toward the trusted person. But further details tend to describe the trusted person rather than the state of mind or dispositions of the person who trusts. Govier (1997, 4), to take an example from the philosophical literature writes: “To trust a friend is to believe that her motivations (toward ourselves) emerge from affection, care, and concern, and not from dislike, ambition, or egoism.” To take an example from the risk literature: Peters et al (1997, 43f) consistently describe trust in terms of the truster’s ‘perception’ of such factors as ‘knowledge and expertise; openness and honesty; and concern and care’. This phenomenon, of confusing trust with trustworthiness, has been pointed out by Russell Hardin (1996):

Many discussion of trust run trust and trustworthiness together, with claims about trust that might well apply to trustworthiness but that seem off the mark for trust. In particular, writers often transfer to trust the moral approbation that might be thought applicable to trustworthiness. (Hardin, 1996, 28)
It seems that any account of trust at least conceptually should attempt to keep the following separate: (i) what is true of the truster (in terms of behaviour, mental states, beliefs, and attitudes) when trusting; (ii) what true about the trusted other, if trustworthy; and (iii) the reasons for the truster to trust the other (including beliefs about trustworthiness). A description of the truster may include a description of his or her beliefs about the trusted person, but such beliefs need not exhaust the phenomena of trust. Baier argued in a much cited passage that in trust we do not only feel mistaken but betrayed if disappointed (Baier 1986, 240ff), indicating that trust is not only a detached belief about something but entails also a more emotional or non-cognitive involvement from the truster. In particular, if we want to include the kind of trust that a baby has for a parent, as Baier (1986) does, or the kind of trust that a young pupil has towards a teacher that Hertzberg discusses (1988), this seem to include other aspects not exhausted by a referral to beliefs about trustworthiness.

Furthermore, not distinguishing between (i), (ii), and (iii) excludes the possibility of exploring the relationship between beliefs about trustworthiness and reasons to trust. This excludes the possibility of trusting in spite of reasons not to, trusting without any beliefs about the truster, and refraining from trusting while believing in the trustworthiness of the other. Judith Baker (1987) pointed to the tendency to trust someone also against evidence. The confusion of reasons to trust; trust; and beliefs about trustworthiness are particularly apparent in empirical studies of trust, where trust is often investigated by asking the participants in a survey if they perceive a particular source as trustworthy (see e.g. Poortinga and Pidgeon 2003). Siegrist and colleagues (Siegrist et al 2003; Earle and Siegrist 2006) include the reasons for trusting in the definitions of their two trust concepts: ‘Trust’ is, in their account, a willingness to make oneself vulnerable to the other based on perceived value similarity. ‘Confidence’ is a belief in future events occurring as expected based on knowledge of past performance. In this way the rationale that motivates trust is confused with trust as such and, for instance, the possibility of trusting someone whom the truster perceives as committed to dissimilar values are excluded by definition (see Essay V).

In my own account (Essay III) I have defined trust as ”comfortable defocusing”. Trust is then characterized partly as a mental disposition to defocus from thoughts or
concern about whatever is in the hands of the trusted combined with a positive feeling towards the trusted other or concerning the object that trust concerns. Such an account of trust consequently says very little about what the other is supposed to do in order to be trustworthy and the reasons for when trust is rational since this is something the truster defocuses from. I have therefore (Essay IV) sketched an outline for necessary properties of trustworthiness that do not depend on what the truster expects from the trusted. Trustworthiness as I have defined it is to live up to social norms for expectations similar to Mullin’s ideas (Mullin 2005), together with a sensitivity for the expectations of the truster, similar to Jones’s and Pettit’s notion of trust-responsiveness (Jones 1996; Pettit 1995), and commitment in motivations over time, as implied in virtue ethical accounts of trustworthiness (Potter 2002). The idea is then that a person can be trustworthy without being trusted by anyone, and that a person can trust another without having a particular belief about the trustworthiness of the other. This lets us explain the phenomena referred to as ‘therapeutic trust’, when we entrust something to another person, who is not yet trustworthy, in order to encourage trustworthiness (Horsburgh 1960).

It has been pointed out by many writers that trust implies vulnerability or exposure to harm (see e.g. Baier 1986, 1992a; Rousseau et al 1998). In this thesis it is proposed that vulnerability is partly an effect of trust as comfortable defocusing. The vulnerability dimension of trust is the result of being less attentive and cautious than toward something we are actively focused upon. This vulnerability is in a way a risk taking necessary for allowing others to take care of things for us. To the extent that the truster is more competent, careful and conscientious than oneself or alternative trustees, this vulnerability may only be instrumental to greater safety (although less personal control). Trust, is thus, often referred to as “accepted vulnerability” (see Baier 1986, Rousseau et al 1998). This vulnerability can also be further enhanced by the unequal relationship between decision makers and public in terms of influence and access to information (see Essay II). The risk taking involved in trust seems then to depend more on the risks involved in the situation than on trust. The truster is however made more vulnerable to those risks when also trusting (see Essay III).
The reasonability of trust does thus not depend on expected trustworthiness alone; in order to refrain from trust the alternatives need to be considered. The alternatives, besides distrust, that would come into question are stances such as: indifference, suspicious reliance, protest and control. What these stances have in common is that they cannot be characterized as comfortable defocusing, that is, they are either actively focused stances directing attention towards the management of risks (e.g. control or protest) or they are defocused but not decidedly comfortable (e.g. suspicion or indifference). The common approach to dealing with the vulnerability of trust, most apparent in ‘blind’ or ‘naïve’ trust, is to recommend a moderate kind of trust, variously named ‘critical trust’, ‘reasonable trust’, ‘mature trust’, etc (see McGeer 2002; O’Neill 2002b; Pidgeon et al 2007). In effect it is trust with a suspicious or watchful eye on that which has been entrusted. Pettit (1998) suggests that citizens should adopt an attitude of trust, but act as if they did not trust, toward decision makers in a state. In my account this would not be trust at all and not provide any of the positive sides to trust, such as the truster not having to worry about that which is entrusted and being able to fully direct her mind to other matters. However, the balance between too much and too little trust does not have to be solved on an individual level. Instead of combining the contrasting attitudes of suspicion and trust into an ambivalent composite of both, distrust can to a large extent be delegated. Most of us have neither time for active distrust nor control, but if these stances are left to trusted others, we can then trust, knowing that someone else keeps out a suspicious eye for us.

The best motivated trust for potentially hazardous industries would probably be one that thrives in an environment of a distribution of both trust and control. That is, the different stances could be shared among different individuals in different positions in society. An example of such a combination of different attitudes and approaches directed to a risk agent could be: regulations and licences restricting risky activities (control); governmental agency surveying emergency planning and safety issues (suspicious reliance), environmental activist scrutinizing the activities (protest), and

---

3 Pettit (1998, 310-311): “People may have an attitude of personal trust because they believe that the authorities are uncorrupt and that they will reliably behave in the proper manner. But there are good reasons, nonetheless, why they may behave as if they had an attitude of distrust, insisting on the necessity of various checks and constraints”.
possibly even improvement proposals from a group of active citizens (active involvement), all watching or related to the same risk responsible organization. Such a distribution would make way for a general trust in terms of feeling comfortable about being defocused among the majority of people since trust relies not only on the trustworthiness of the plant management but on a wider group of people. Control is then delegated to another trusted party keeping a watchful eye on the decisions, which hopefully would make it safer for others to trust. Hence a general recommendation to trust or not to trust is to ignore precisely the kind of delegation of attitudes and measures that could make a general societal trust both reasonable and possible. It is from this perspective that trust in risk communication cannot be treated without its social context, and it seems that the sociological concepts of ‘system trust’ (see e.g. Jalava 2003), and the concept of ‘epistemic trust’ (see e.g. Sjöberg & Wester Herber 2007) have more to contribute here than that of reasoned or rational trust (see Essay V).

In the case of individual risk taking the source of vulnerability is not so much the decisions of others but actions or behaviour of the individual. Refraining from taking risks presumes that the person is in a position of power to make less risky choices (Essay I). Trust and risk communication will have little effect on individual risk taking unless there are real options for concerned individuals. When communicating risks to a wider group we must also assess the influence of that group. Vulnerability is increased if an individual has few options, or small means to counter-act negative outcomes. The risk from relying on the word of someone is greater if there is no way to check what is said. Thus the normative issues of trust and distrust cannot be addressed without also addressing the influence of those expected to trust.
Essay I

Out of the Ashes:
Hope and Vulnerability as Explanatory Factors in Individual Risk-Taking

Essay I is a comment on risk perception theory, particularly on the psychometric model. In the risk perception literature individual risk-taking is, implicitly or explicitly, presumed to be a result of either valuing the benefits from risk taking or a failure of comprehending the severity or probability of risks. Thus individual risk taking can be counter-acted through risk communication about the nature of such risks. The essay challenges this view, and points to a kind of risk that does not seem to fit this explanation. These risks are referred to as risks from vulnerability. Examples of such risks would be: dangerous means for emigrating; amateur abortions and compensated kidney donations for repaying debts. A hypothesis intended to complement prevalent explanations to risk taking is proposed: some risks taken could be explained in terms of the poor outset conditions of the risk taking individual; a lack of reasonable alternative actions to choose from; the hope that an option is believed to offer; and the liability to disinformation concerning alternative actions and their risks. In difficult cases there is a combination of these elements: the individual risk taker may be offered a hope inspiring an escape, but a risky one, out of an already unbearable situation, with no other reasonable alternatives in sight, and the only source of information as to the risks may be someone earning his or her living in offering such escapes. The main conclusion is that these risks cannot be solved through traditional risk communication without also providing information on substantial alternatives.
Essay II

Asymmetry in Risk Communication

Essay II addresses the democratic approach to risk communication and the ideal of risk communication as a dialogue between more or less equal stakeholders. Risk communication is often framed in the literature as being either technical and one-way or democratic and two-way. In this context democratic and two-way communication is often preferred either as an ideal for building mutual trust between, for example, an industry and the local population, or two-way communication is regarded as a complementary or a more developed form of communication. The essay starts from these democratic approaches and distinguishes between two-way communication and a democratic approach, although both are contrasted to one-way dissemination of information to the public. The essay argues that even the democratic or two-way approaches to communication are often in practice far from the stated ideal of equal participants exchanging ideas. Often the initiative for the dialogue, the background knowledge, and the influence over the communication process and the risk inducing activity are asymmetric. The essay identifies three different levels of influence in the communication process and consequently three different levels of influential asymmetry. These levels are: communicative influence (over the communication process); informative influence (over access to information) and risk role asymmetry (over risk decisions). These aspects can strengthen or counteract the overall balance in influence between the parties depending on how they are distributed. A typology of different kinds of risk communication practices is developed based on such distributions. Seven kinds of risk communication practices are presented, fitting into the categories of symmetrical communication; communication from scattered asymmetry; and communication from clustered asymmetry. It is concluded that the practices most often discussed in the literature sort under the category clustered asymmetry, with communicative influence, information, and risk role influence tilting to one party’s advantage.
Essay III

Trust as Comfortable Defocusing

Essay III starts from an idea of trust as comfortably defocusing from that which we trust another person to do. This idea is spelled out against previous ideas of trust as positive expectations and trust as the lack of distrust. The essay proposes a definition of trust in an attempt at combining elements from both lines of thought while avoiding the negative conclusions of trust not being a first person mental state. It is argued that trust, can be characterised as a first-person disposition to comfortably defocus and in this sense both imply expectations about the other and a lack of attention and reflection upon that which in entrusted to someone else.

The idea of comfortable defocusing is insufficient to explain trust by itself. Two additional conditions for trust are added to the account. First, it is argued that the disposition to defocus needs to be based on a belief that the trustee will take care of that which is entrusted. If such a belief were not the reason for comfortable defocusing, a distracted mind or ignorance would also qualify as trust. Second, it is argued that what the trusted person is expected to do or care for must be of concern to the truster. The belief that another person will take care of those things that he or she routinely can be expected to care for would not qualify as trust unless this is of concern to the trusted. It is argued that in trust, it is the belief that the trusted person will care of a certain matter that is the primary cause for the truster to be comfortably defocused, and thus, if this belief is disrupted the trusted would no longer be able to defocus comfortably but attend to the matter or feel uncomfortable about the it. My account of trust should be understood as a three-part explanation consisting of: (i) the truster’s comfortable defocusing from the trusted person’s part of caring for X, based on a (ii) belief about the trusted person’s care for X, concerning (iii) something of concern for the truster.
Essay IV

Morally Justified Untrustworthiness

This essay departs from an idea brought forward by Annette Baier that it sometimes can be morally right to betray trust. This essay examines this idea from two different approaches to the notion of trustworthiness. In the first approach trust defines trustworthiness; to be trustworthy is to be what the truster believes or relies on one to be. On this reading a person is trustworthy to the extent that she delivers what the truster trust asks from her. I refer to this idea as the fulfilment of expectations view of trustworthiness. Righteous untrustworthiness is then simply a matter of refraining, for good moral reasons, from complying with these expectations. If A trusts B to do the immoral act X, then trustworthiness demands that B does X, but morality that he doesn’t.

An alternative approach to trustworthiness is to refer to character or dispositions. To be trustworthy is then to be the kind of person that is suitable for trust regarding a certain matter, should someone choose to. Trustworthiness is then a reliable trait of character, regardless of whether one is trusted or not. I refer to this as character based trustworthiness. The virtue ethical versions of character-based accounts are however difficult to combine with an idea of a morally justified untrustworthiness, since being trustworthy is to be a virtuous person.

In this essay it is argued that two aspects are essential for every reasonable notion of trustworthiness: reliability in terms of commitment to trust and predictability in terms of compliance with norms. Together these two aspects can explain the value generally placed on trustworthiness and how a trustworthy person allows for the truster to fully rely on her. In order to be untrustworthy in a morally justified way two major conditions need to be fulfilled. First, what is expected from the trustee needs to be either immoral or part of an ongoing immoral or suppressive relationship between truster and trustee, or tie the trustee to social expectations of a suppressive kind. Second, the trustee must not be in a position to explicitly reject trust.
Essay V

Un-Confusing the Notion of Trust

Essay V returns to the notion of trust and risk communication. Few concepts have received so much attention within risk research as that of trust, but two decades after it was first brought to attention there is still no unified understanding of it within the discipline. The lack of a theoretical framework of trust for risk communication purposes has generated an abundance of competing conceptual proposals borrowing from other disciplines resulting in a conceptual chaos that is repeatedly recognized as problem within the field.

A common approach to clarify a concept is to distinguish it from other closely related concepts and spell out the differences. In this essay two of the most influential trust-distinctions from two different disciplines are discussed and compared. The first of the two is the trust and confidence distinction that has become influential in risk research, originally developed by Niklas Luhmann, but later adopted for the risk context by Timothy Earle, Michael Siegrist and Heinz Gutscher. The second distinction is the trust and reliance distinction that has been very influential in the philosophical literature on trust, primarily since Annette Baier pointed to it in her seminal account of trust as reliance on the goodwill of another. Both set of notions are presented and assessed against six basic criteria for fruitful and consistent distinctions. These criteria are: same category, distinguishing property, relevance to discourse, mutual exclusiveness, conceptual economy, and exhaustiveness.

It is concluded that the proposed distinction between trust and confidence has been employed to cover a whole range of different distinction that fail to make it clarifying in terms of mutual exclusiveness, that makes the over-arching category elusive, and hence makes it difficult to employ in empirical research. However, several of these implied sub-distinctions taken by one by one are both relevant and mutually excluding conceptually. In particular the distinction between system trust and person trust seem promising.
Furthermore it is concluded that the distinction between trust and reliance is problematic as a distinction, particularly since there are at least two different concepts of reliance implied. The idea of trust as a subcategory of reliance however seems to add crucial parts to other explanations of trust in terms of merely beliefs and expectations.

References


