Thick Concepts in Practice:
Normative Aspects of Risk and Safety

Niklas Möller

STOCKHOLM 2009
This doctoral thesis consists of the following essays preceded by an introduction:

I. Möller N. “Thick Concepts and Practice: Strengthening the Non-Naturalist Case”, submitted manuscript.

II. Möller N. “Moral kinds, Natural Kinds and the Open Question Argument”, submitted manuscript.


Essays IV-VI are used with kind permission from Blackwell Publishing (Essay IV), Inderscience (Essay V, DOI 10.1504/IJRAM.2009.023154) and Elsevier (Essay VI).


Niklas Möller, Division of Philosophy, Department of Philosophy and the History of Technology, Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden.
© Niklas Möller 2009
ISSN 1650-8831
Printed by US AB, Stockholm, Sweden, 2009
To Eva
Abstract

The thesis aims at analyzing the concepts of risk and safety as well as the class of concepts to which they belong, thick concepts, focusing in particular on the normative aspects involved.

Essay I analyzes thick concepts, i.e. concepts such as cruelty and kindness that seem to combine descriptive and evaluative features. The traditional account, in which thick concepts are analyzed as the conjunction of a factual description and an evaluation, is criticized. Instead, it is argued that the descriptive and evaluative aspects must be understood as a whole. Furthermore, it is argued that the two main worries evoked against non-naturalism – that non-naturalism cannot account for disagreement and that it is not genuinely explanatory – can be met.

Essay II investigates the utilization of the Kripke/Putnam causal theory of reference in relation to the Open Question Argument. It is argued that the Open Question Argument suitably interpreted provides prima facie evidence against the claim that moral kinds are natural kinds, and that the causal theory, as interpreted by leading naturalist defenders, actually underscores this conclusion.

Essay III utilizes the interpretation of the Open Question Argument argued for in the previous essay in order to argue against naturalistic reduction of risk, i.e. reduction of risk into natural concepts such as probability and harm. Three different normative aspects of risk and safety are put forward – epistemic uncertainty, distributive normativity and border normativity – and it is argued that these normative aspects cannot be reduced to a natural measure.

Essay IV provides a conceptual analysis of safety in the context of societal decision-making, and argues for a notion that explicitly includes epistemic uncertainty, the degree to which we are uncertain of our knowledge of the situation at hand. Some formal versions of a comparative safety concept are also proposed.

Essay V puts forward a normative critique against a common argument, viz. the claim that the public should follow the experts’ advice in recommending an activity whenever the experts have the best knowledge of the risk involved. The importance of safety in risk acceptance together with considerations from epistemic uncertainty makes the claim incorrect even after including plausible limitations to exclude ‘external’ considerations. Furthermore, it is shown that the scope of the objection covers risk assessment as well as risk management.

Essay VI provides a systematized account of safety engineering practices that clarifies their relation to the goal of safety engineering, namely to increase safety. A list of 24 principles referred to in the literature of safety engineering is provided, divided into four major categories. It is argued that important aspects of these methods can be better understood with the help of the distinction between risk and uncertainty, in addition to the common distinction between risk and probability.

Keywords: thick concepts, non-naturalism, open question argument, risk analysis, safety, epistemic uncertainty, values in risk assessment, safety engineering

Division of Philosophy, Department of Philosophy and the History of Technology, KTH, Sweden.
Acknowledgements

I wish to express my gratitude to all those who have given me advice and support in the making of this thesis. First of all, I am very grateful to my supervisors Sven Ove Hansson and Martin Peterson for their valuable comments, suggestions and support without which this thesis would not have come about. Secondly, I am grateful to all my colleagues at the Division of Philosophy who have contributed with valuable suggestions and comments on earlier versions of the papers. Special thanks go to Kalle Grill and Lars Lindblom for their substantial comments and lengthy discussions of all conceivable philosophical topics; Per Wikman-Svahn and Anders Persson for their mental and well as physical heavy lifting support; and John Cantwell for his constructive criticism on the later papers of the thesis.

During my semester at Georgetown in the fall of 2007, I was supervised by Mark Lance and Maggie Little. My innumerable discussions with Mark in his crowded office broadened my philosophical outlook, and my Tuesday lunches with Maggie – with their intertwining discussions of philosophy and life – are up there in whatever proximity to Platonic Heaven a pragmatist atheist such as myself may construe. For this, as well as for their continued support, I am immensely grateful.

Finally, I want to thank four persons for their particular contributions to this thesis. Eva Erman, for her love, her constructive criticism and never-ending deliberation, but also for her decisive role in the coming-about of this thesis: It was watching the joy she emitted when engaging in her own PhD studies – combined with her unrivalled talent for making things happen – that helped put me on this track. Lea and Elias, who (ten minutes apart) entered the scene during the course of this thesis and helped me – in their inescapably distinct way – to keep things in perspective. And Britt Erman, for her careful proofreading and the deeply supportive mode in which she always returned the corrected papers.

This work has been financially supported by Krisberedskapsmyndigheten, The Swedish Emergency Management Agency. Furthermore, STINT – The Swedish Foundation for International Cooperation in Research and Higher Education – financed my semester at Georgetown University, Washington DC. This support is gratefully acknowledged.

Stockholm, April 2009
Niklas Möller
Contents

Abstract iv
Acknowledgements v
Introduction 1
  1. Central claims 3
  2. Theoretical background 7
  3. Looking forward 18
  4. Preview of essays I-VI 19
References 23

Essay I 27
Möller N. “Thick Concepts and Practice: Strengthening the Non-Naturalist Case”, submitted manuscript.

Essay II 55
Möller N. “Moral kinds, Natural Kinds and the Open Question Argument”, submitted manuscript.

Essay III 81

Essay IV 111

Essay V 133

Essay VI 159
Introduction

Risk research is a discipline in rapid development with contributors from many areas of the natural and social sciences. This reflects a growing concern about risks in society. Both professional and non-professional awareness of risks is increasing, and much effort is put into risk assessment, risk management and risk communication. As a consequence, there are now well-developed societal practices in place involving risk and safety. Still, the central concepts of risk and safety remain unclear. When characterized, risk and safety are often either treated as relatively straight-forward natural science concepts, or, to the contrary, as fundamentally subjective notions ill fitting for scientific study. More often still, the terms of ‘risk’ and ‘safety’ are used merely to mark the domain of study, and a more precise meaning of the concepts is left unclear.¹ However, without an in-depth understanding of its central concepts, the subject matter of risk and safety research remains fuzzy and it is unclear what the objective of reducing risk and achieving safety really amounts to.

In the present thesis, the risk and safety concepts provide a focal-point of two debates: the debate in applied philosophy and risk research about how to understand the risk and safety concepts, and the debate in metaethics about how the important class of ‘thick concepts’ should be analyzed. A guiding conviction has been that the corresponding areas of inquiry may cross-fertilize each other. Metaethics deals with the status of normative concepts, and insights from this domain may enhance our understanding of risk and safety – since these, as will become clear, are normative concepts. Risk and safety research and practice, on the other hand, provide a rich background for analyzing thick concepts.

The investigation in the current thesis is thus performed on two interrelated levels. The investigation of thick concepts is performed on what may be called the meta-theoretical level. Thick concepts are, roughly, concepts that have both descriptive and evaluate features, and therefore form a “middle” class between paradigmatic natural science

¹ This is the common usage in regulation and procedural documents. Safety of Nuclear Power Plants (IAEA 2000) is a typical example. Here, focus is on describing rules and procedures to enhance safety, but any direct and precise characterizations of the concepts of risk and safety are missing.
concepts such as photons and gold and paradigmatic evaluative concepts such as rightness and fairness. On the applied level, the investigation focuses on the concepts of risk and safety. These concepts are important for decision-making in general and central (as the very name of the field suggests) for risk and safety research. Furthermore, risk and safety clearly display the descriptive/evaluative duality crucial for thick concepts. On the one hand, risk and safety are central concepts in the scientific enterprise of risk assessment. On the other, they have evaluative force, i.e. that an action carries a high risk is, prima facie, a reason against it.

Broadly speaking, there seem to be two largely disparate views of the central concepts of risk and safety. On the one hand, natural scientists tend to perceive risks as natural science phenomena, as properties in the world independent of individual beliefs: either the bridge is safe to pass or it is not. Social scientists, on the other hand, often claim that risk is something essentially subjective or socially constructed: people from different cultures as well as within a culture have very different views on what constitutes a risk and how severe a risk may be, and there is no fact of the matter over and above these individual or cultural views. Paul Slovic, for example, notes that defining what constitutes a risk is an exercise of power, and concludes that there is no such thing as “real risk”. The various conceptions of risk and safety are often implicitly given, or even merely hinted at, and a thorough justification is seldom provided.

A central hypothesis of the thesis is that there is a better way of understanding the concepts of risk and safety, a ‘middle way’ emphasizing that they are normative, which at the same time does not entail a constructivist denial of ‘real risk’. Exploring the notions of risk and safety in more depth reveals, it is claimed, several important aspects of normativity. Particular emphasis is here put on the concept of safety, which is especially under-theorized.

For the meta-theoretical part of the thesis, the resulting normative view on risk and safety is employed in investigating the notion of thick

---

3 This belief is influenced by contemporary socio-pragmatist views in philosophy of language and metaethics; in particular, the views of Robert Brandom (1994) and John McDowell (1994, 1998) who have both been influential to my perspective, if in a basic, background sense that has not always made it to the reference section.
concepts themselves. Thick concepts have been treated foremost by contemporary moral philosophers, due to the ‘middle position’ they occupy between descriptive and evaluative concepts. Understanding how these concepts should be viewed arguably provides a key for understanding ‘thin’ moral concepts (*right, good*, etc) as well – consequently, thick concepts are often alluded to in metaethical arguments from all camps, realists as well as antirealists of different fractions.⁴

Traditionally, theorists interested in the role and nature of thick concepts have focused on every-day concepts such as *courage* and *cruelty*. While there certainly are lessons to be learnt from arm-chair analyses of concepts exclusively from the vernacular, the concepts of risk and safety have the benefit of also figuring in greater social and scientific contexts. Hence, there is a well-developed practice in place for risk and safety, which enables us to reach beyond mere arm-chair philosophizing when studying how these concepts function. The articles in the current thesis thus have the double aim of supplying a better understanding of the very concepts of risk and safety, as well as using this knowledge for elucidating the nature of thick concepts in general.

1. **Central claims**

As a consequence of the twofold ambition of the thesis of providing some insights into two areas of philosophy – applied philosophy as well as what typically goes under the umbrella category of metaethics – its corresponding claims are made on two levels. On the meta-theoretical level, the thesis argues for an analysis of thick concepts in which the descriptive and evaluative aspects of the concept are necessarily intertwined (in a sense that will be further explained). On the more applied level, the thesis argues for different senses in which the concepts of risk and safety are normative – i.e. includes non-descriptive, evaluative features – and for the further claim that this normativity is irreducible, i.e. cannot be expressed using only non-normative terms. Scientific conceptualizations of risk and safety using only non-evaluative concepts are thus insufficient for capturing the phenomena of risk and safety.

Starting on the applied level, a central claim in the investigation of the concept of safety is that the notion of *epistemic uncertainty* should be included in the analysis. In the traditional scientific analysis of the concept of risk (in the sense of interest here), it is analyzed as the expected value of harm, which is the probability weighted sum of the severities of the potential harmful events. In many ways this conception makes perfect sense: the more severe a harmful event and the larger the probability that this event will take place, the higher the risk. This is certainly a central intuition about the concept. Moreover, safety is traditionally defined as the antonym of risk: the lower the risk, the higher the safety. I argue, however, that the notion of epistemic uncertainty must also be included in the concept of safety. The reason is that an expected value may reflect very different epistemic situations in a way relevant for safety. For example, let us assume that two means of transportation have the same expected value of harm, according to the best expert judgment. The former means of transportation, however, utilizes a new design, never used before in actual conditions, whereas the latter utilizes a traditional design for which there is extensive statistical data of performance, reliability, accidents etc. Even if the new design has the same presumed expected values of harm, we may claim that the traditional design is safer due to the fact that the certainty with which we may trust its value is higher – it has, that is, lower *epistemic uncertainty*. For the new design, there is thus a greater danger that the probabilities of some harmful events are misjudged, and this is what motivates our preference – in terms of safety – for the traditional design in this case.\(^5\)

I argue that epistemic uncertainty is an evaluative aspect of the concept of risk and safety,\(^6\) over and above the scientific underpinnings of the expected value itself. There is no consensus as to how this notion may be described in detail, although there are several competing models and suggestions including second-order probability and fuzzy set theory. It is further argued that there are at least two other senses in which risk

\(^5\) Epistemic uncertainty is also evident in evaluating the harmful events, but arguably to a lesser extent in many actual safety cases.

\(^6\) In *Essay IV* and *Essay V*, I use the traditional probabilistic (expected value) notion of risk as a point of departure – and critique – in my analysis of safety. However, the same epistemic uncertainty argument may naturally be used to argue against the probabilistic notion of risk as well. Indeed, that is what I choose to do in *Essay I* and *Essay III* when focusing on the normative status of both concepts.
and safety are normative: what I call *distributive normativity* and *border normativity*. Distributive normativity is the role that different distributions of potential harm play in risk and safety ascription. In some cases of risk, where the distribution of potential harmful outcomes for, say, a method of power production is uneven to such an extent that some people face much higher potential harm, it may reasonably be claimed that this affects the total risk. Border normativity is what I label the normative aspect of how to delimit which events should count as relevant for assessing the question of the risk and safety in the case at hand. Cases of traffic incidents that on the harm dimension are qualitatively identical (resulting in death, for example) may be qualitatively different in other important aspects that matter for risk and safety ascription. For example, one traffic incident may be due to a suicide attempt, and not merely an unintentional accident. In some cases suicide events should arguably not be considered relevant for traffic safety, whereas there are other circumstances in which they should. Deciding which events should be included in our safety assessment is, I argue, a normative aspect beyond the probabilistic conceptualization.

On the meta-theoretical level, I argue for a specific understanding of the class of concepts to which risk and safety belong, i.e. thick concepts. As mentioned in the previous section – and as will be discussed in more detail in the next section – a thick concept is a kind of concept that has a significant degree of descriptive content and is evaluatively loaded at the same time. On the traditional view, a thick concept may be understood as a descriptive concept such as *length* and *water*, but with an evaluative addition: whereas understanding that X is water is compatible with both a pro-attitude towards it (e.g. if you are thirsty) and a contra-attitude (e.g. if you fall into it and cannot swim), understanding that X is cruel typically involves a negative attitude towards X. ‘Addition’, since on the traditional analysis, the evaluative aspect is merely an add-on in the sense that the descriptive content is sufficient for capturing the extension of the thick concept in question. The general claim of the thesis is a denial of this traditional analysis: it is argued that thick concepts cannot correctly be given such a conjunctive analysis. Thick concepts have, differently put, no distinct ‘natural shape’ – the descriptive and evaluative aspects are intertwined and there are no descriptive delimiting boundaries. This
means that there is no non-evaluative description that alone manages to capture the meaning of a thick concept.\footnote{See the next section for a more extensive treatment.}

In arguing for the meta-theoretical claim about the semantic and epistemic status of thick concepts, I make use of the result of the investigation into the concepts of risk and safety; in this way, there is a close relationship between the applied and the theoretical level of the thesis. I claim that a deeper investigation into the normative aspects of risk and safety reveals the unreasonableness of a naturalist solution to the problem of picking out the correct properties. There is, I argue, no reasonable way of finding necessary and jointly sufficient descriptive conditions for application of a thick concept, and alternative strategies of making sense of the naturalist position remain unconvincing. I argue that this naturalist failure reflects the essential interconnection between the descriptive and evaluative aspects a thinker needs to grasp in order to correctly apply the concept in new circumstances. Furthermore, I argue that a naturalist view has severe problems capturing the difference between explanatorily efficacious conditions and merely background conditions for the thick concept in question.

The thesis also investigates a second route to irreducible normativity, Moore’s Open Question Argument. According to this argument normative concepts such as goodness and justice do not have the same meaning as any natural concept; instead, they are non-natural concepts. In short, Moore’s thought is that whatever natural concept \( N \) we may propose as having the same meaning as a moral concept \( M \), we may always intelligibly ask whether something that is \( N \) really is \( M \) – it is always an ‘open question’. But, he argues, if the meaning of \( M \) and \( N \) were identical, the question should not be open. Hence, the meanings of the moral concept and the natural concept are not the same.\footnote{See the next section for further explanation of the Open Question Argument.}

In the thesis, the Open Question Argument is defended and utilized for two purposes. First, it is used (in accordance with Moore’s intention) to defend moral non-naturalism. The main contemporary naturalist critique of the argument comes from direct reference theories of meaning in general and the causal regulation theory in particular.\footnote{See the section below.} I argue that a
causal regulation theory in fact underscores the conclusion of the Open Question Argument. The naturalist critique depends heavily on the analogue between natural kinds and moral kinds, but the Open Question Argument provides prima facie evidence against the claim that moral kinds are natural kinds, and it is argued that the causal theory, as interpreted by leading defenders of moral naturalism, does not overturn this evidence. Whereas the thinker doubting that something is water if it is H₂O may be presented with overturning theoretical commitments, there are no corresponding, deeply held theoretical commitments to put forward in the moral case.

Secondly, the Open Question Argument is applied to the specific thick concepts of risk and safety, in order to argue for the irreducible normativity of these important concepts. The normative aspects of risk and safety are presented, and it is argued that there are no overruling theoretical commitments that justify naturalistic reductions of risk and safety.

2. Theoretical background
Risk and safety as thick, action-guiding concepts are the main focus of attention in this thesis. Action-guidance is of interest in several philosophical areas, notably moral theory. Metaethics is the sub-discipline of moral theory studying the status of moral claims. Naturalism and non-naturalism in moral and the notion of thick concepts are heavily discussed in contemporary metaethics, and subsection 2.1 provides a short introduction to the central debates.

Whereas moral philosophy mainly deals with the ‘narrow’ question of how to act, viz. how to act morally right, there are other areas of philosophy where action-guidance plays a central part. The growing field of practical reasoning deals with how we should reason about what to do in more general terms. However, decision-making has most systematically been studied in decision theory. Several decision theoretical themes and concepts are relevant for the essays in the thesis – notably the

---

10 Roughly, whereas normative ethics typically focus on the first-order question of which actions are right or wrong, metaethics deals with second-order questions of the moral enterprise – the ontological, epistemological, semantic, and psychological aspects of moral thought and practice.
conceptualization of probability, epistemic uncertainty, and utility, as well as decision rules such as Maximizing Expected Utility – and the subsequent subsection 2.2 introduces the central issues.

2.1. Naturalism and non-naturalism

The term ‘non-naturalism’ is used in the thesis for two related but distinct views, a view in metaethics (in Essay 2) and a view about the analysis of thick concepts (in Essay 1). As the term indicates, these views oppose ‘naturalism’ – but what, more exactly, does this opposition entail?

‘Naturalism’ is a term with many philosophical uses. In a broad sense, it may be said to denote the philosophical view that “everything (objects and events) is a part of nature, an all-encompassing world of space and time”. Moreover, it typically involves the idea that the scientific method is valid in all areas of inquiry. Endorsed in the realm of morality, it entails that moral values, moral obligations, etc also fit into a naturalistic world view. This broad sense of naturalism which is compatible with many metaethical views – e.g. realism, conventionalism and antirealism – is not disputed in the thesis.

The notion of naturalism in focus is rather a more one which is often used in contemporary metaethics. Here, naturalism denotes the ontological, moral realist view that there are objective moral facts and properties, and that these facts and properties are nothing but natural facts and properties. As such, it is opposed to both moral anti-realism (there are no moral facts) and moral non-naturalism (moral facts are not natural facts). Closely related to ontological moral naturalism is semantic moral naturalism, claiming that moral concepts can be analyzed in non-normative terms. Typically, the semantic moral naturalist claims that a moral concept such as good may be analyzed as a natural concept, e.g. maximizes the amount of well-being. A corresponding ontological

---

12 Mautner 1996.
15 Realism is the view that moral facts and properties are objective facts and properties of the world; conventionalism the relativist view that ethical right and wrong are fixed by (social) conventions; and anti-realism the view that there are no moral facts – rather, the function of moral discourse is to express non-cognitive states such as emotions, pro/contra-attitudes etc.
commitment is then that a moral fact, such as the fact that it is right, in situation S, to show kindness to person P, is true due to the fact that showing kindness to P in S maximizes well-being.

When addressing moral naturalism and non-naturalism directly, these semantic and metaphysical views are intended. Likewise, it is in analogy with these views that I distinguish between naturalist and non-naturalist views of thick concepts.

2.1.1. The Open Question Argument
The starting point of the moral naturalism/non-naturalism debate is G.E. Moore’s seminal Principia Ethica (1903), in which he argued against the “naturalistic” ethical views of philosophers such as Jeremy Bentham, Herbert Spencer and T.H. Green. Moore put forward an argument attempting to show that moral properties are non-natural properties, an argument which came to be known as the ‘Open Question Argument’. Let us assume that ‘M’ is a term expressing a moral property, and ‘N’ a term expressing a natural property. Moore’s argument is that ‘N’ and ‘M’ cannot be identical in meaning, whatever natural and moral terms we choose, since the following question is always open: ‘x is M, but is x N?’ By ‘open’, Moore means that a competent user of M and N may reasonably doubt the answer to the question. On the contrary, a competent user of, say, the concepts of brother, male and sibling cannot reasonably doubt the answer to the question ‘Sam is my male sibling, but is Sam my brother?’ If she doubts the answer, she lacks understanding of at least one of the concepts in question.

Focusing on the concept of goodness, Moore tested his question on some of the major naturalistic candidates of the time – such as whether good is what we desire to desire, or perhaps what is maximally conducive to human welfare – and argued that whatever natural property selected, the question is always open. He concluded that moral terms are

---

17 Moore 1903.
18 Naturally, this assumes the standard interpretation of the terms involved. In some contexts, ‘brother’ may perhaps be given unorthodox interpretations such that the question becomes open. I may perhaps claim that Sam is my male sibling but not my brother if he is “disowned”, or if ‘brother’ is interpreted as gender role and ‘male’ as biological sex.
19 Note that it is controversial whether we can understand concepts such as welfare and well-being naturalistically. Cf. Railton 2003 for arguments in the affirmative.
indefinable in natural terms, and thus that moral properties are non-natural properties.

Moore’s Open Question Argument convinced many of his contemporaries of the correctness of non-naturalism. Later, the argument has been used in one form or the other to argue for different versions of anti-realism.\textsuperscript{20} With the help of recent developments in philosophy of language, however, naturalists have heavily criticized the argument. A moral property, the objection goes, may be a natural property even though it is perfectly reasonable for a competent language user to doubt any suggested natural property identification – in other words, property identification is not generally \textit{a priori} but \textit{a posteriori}, it requires investigation into how things stand in the world.

In its current form, the naturalist objection has been based on the Kripke/Putnam causal regulation theory of reference for names and natural kinds.\textsuperscript{21} According to the causal regulation theory, a name or a natural kind term refers through a complex interplay of causal connections between utterances of the term and the corresponding object or kind. Paradigmatically, the natural kind term ‘water’ refers to (and only to) entities with the molecular structure of $\text{H}_2\text{O}$, since it is a substance of the kind that causally has regulated our usage of the term. Moral kind terms function in analogy with such natural kind terms, naturalists argue. Hence, moral kind terms such as ‘good’ or ‘just’ refer to the kind of entities that causally regulate our usage of these moral terms, whether or not this fact is epistemically transparent to us. In other words, just as ‘X is $\text{H}_2\text{O}$, but is it water?’ is in fact ‘closed’, even though it may seem open for most competent language users (since most people arguably do not know enough chemistry), the same may be true also of ‘X would amount to the largest sum of well-being, but is it good?’, regardless of whether it seems open or not.

This naturalist objection has in turn been under fire, in particular in a series of famous papers by Terence Horgan and Mark Timmons, and the debate is ongoing.\textsuperscript{22} In the present thesis, a case is made for an

\textsuperscript{20} Eg. Stevenson 1944, 1963 and Hare 1952, and later Blackburn 1993 and Gibbard 1990. Cf. also note 15.


\textsuperscript{22} Horgan and Timmons 1990-91, 1992a, 1992b.
interpretation of the Open Question Argument that, contrary to Moore, acknowledges the possibility of a posteriori property identification, but defends the soundness of the Open Question Argument as an argument to the best explanation.\textsuperscript{23} Hence, knowledge of identity of properties is not generally \textit{a priori} but may require empirical investigation. Still, if the Open Question Argument is effective – if the question appears to be open even after due consideration – this suggests non-naturalism and shifts the burden of proof to the naturalist camp.

2.1.2. Thick concepts

Giving a precise characterization of a natural property is not an easy task, and in the thesis I follow the established tradition of deferring to the properties used to describe scientific phenomena. Moore himself offered a number of characterizations to this effect: properties that are the subject matter of natural science and psychology; properties that can be known by means of empirical observation and induction; and properties that exist by themselves in time.\textsuperscript{24} Although the last suggestion is generally considered as implausible,\textsuperscript{25} reference to the scientific enterprise is still the dominating characterization. Natural properties are thus conceived as the properties invoked in scientific explanations or, more narrowly, properties figuring in the laws of nature.\textsuperscript{26}

In the thesis, I focus primarily on the conceptual side of the natural/non-natural divide, and rely on the availability of a descriptive, non-evaluative characterization of certain concepts. If there is such a characterization available for a concept, I treat it as a natural concept and

\textsuperscript{23} Cf. Ball 1988, 1991 for a similar interpretation.

\textsuperscript{24} Moore 1903, §25-27.

\textsuperscript{25} Few properties seem to be able to exist in time by themselves: while a table may be flat and have a round surface, it is indeed hard to understand how this roundness and flatness may exist in time by itself (cf. e.g. Broad 1942). Consequently, Moore later gave up this way of characterizing natural properties (Moore 1942).

\textsuperscript{26} Little 1994, 226 and Vallentyre 1998. The latter characterization in particular is in line with traditional understanding of natural kinds: the close relation between natural kinds and lawlike regularities is emphasized by most natural kind theorists. (Cf. Lange 2007, Boyd 1991, Hacking 1991.) Wolf 2002, 87, puts the common denominator thus: “I take it to be uncontroversial that the distinctive practical task that cannot proceed without natural kind terms is the explanation of natural phenomena. In explaining how and why the world is how it is and does what it does, we take up commitments to some set of [natural kinds] (however many our theories require) and to claims that they behave in certain ways with a lawlike regularity.” Cf. \textit{Essay II} for a discussion of natural kinds and moral kinds.
thus label the corresponding property or kind picked out by this concept as natural. Paradigmatic natural concepts are \textit{water}, \textit{gold} and \textit{tiger} (corresponding to natural kinds) and \textit{length}, \textit{mass} and \textit{redness} (corresponding to properties of objects).

Natural concepts are typically contrasted with evaluative concepts, concepts used to express evaluative claims. \textit{Good}, \textit{right} and \textit{fair} are paradigmatic examples of such concepts. Evaluative concepts are typically \textit{action-guiding}: that an action (or thing) is good provides us with a reason to perform it (use it/buy it, etc). One way of putting this is that these concepts express \textit{values} rather than \textit{facts}: whereas it is matter of fact whether a statue is made of copper, it is a matter of value whether it is a good statue. Similarly, whether the cat is being put on fire is a factual question, but whether this is morally wrong is an evaluative question (of course, moral realists claim that there are moral facts as well – true value statements just express a specific sort of facts).

The paradigmatic evaluative concepts \textit{good}, \textit{right} and \textit{fair} seem to have very little, if any, descriptive content. For something to be water, it has to consist of H\textsubscript{2}O. Nothing similar seems to hold for the paradigmatic evaluative concepts: actions and things wholly different from a natural, descriptive point of view may all be good or right. There are, however, concepts that seem to combine the descriptive features of the natural concepts with the evaluative features of the paradigmatic evaluative concepts (good etc): \textit{courage}, \textit{cruelty}, \textit{kindness}, \textit{loyalty} etc. Such concepts, it seems, have a significant degree of descriptive content: their application is “determined by what the world is like” as Bernard Williams puts it. Yet at the same time their application involves a valuation of, say, an action situation and provides a reason for or against performing it. That an action is kind is, \textit{prima facie}, a reason for performing it – although there

\begin{flushleft}
\footnotesize
\textsuperscript{27} Note that this means that the natural/moral distinction is not a dichotomy, at least not on the outside: a natural concept and a non-natural concept may pick out the same properties. Indeed, the naturalist project, on the interpretation mostly used in this thesis, is to claim that they do.
\textsuperscript{28} The imagined burning of cats in contemporary moral theory is due to Harman 1977, in which he uses the example of a group of children putting a cat on fire in his argument against moral realism.
\textsuperscript{29} Again, whether this is actually so is a matter of substantial moral theory.
\textsuperscript{30} Williams 1985, 129 (and to the same effect on page 141).
\end{flushleft}
may in the present case be other, stronger reasons for some other action (even, perhaps, one that happens to be unkind).

Thick concepts thus seem to combine the descriptive features of natural concepts such as water with an evaluative content similar to the (thin) evaluative concepts such as good and right. How are we to understand this ‘combination’? Many theorists treat it as a conjunctive: a thick concept should be analyzed as a conjunction of a descriptive part and an evaluative part, which, at least in principle may be separated.\(^{31}\) A basic feature of this analysis is thus that the descriptive content of a thick concept may be given in absence of the evaluative content. The expression ‘...is courageous’ could therefore be analyzed as something along the lines of ‘...intended to act in the face of danger to promote a valued end’ and ‘this is (prima facie) good-making’.\(^{32}\) The evaluative part, on this view, may thus be characterized as a ‘prescriptive flag’ attached to the concept.\(^{33}\) In Essay I, I call this view on thick concepts naturalist, since the descriptive part is sufficiently autonomous for picking out the actions etc that sort under the concept, and thus similar to how a natural description presumably may pick out the same property as a moral description. It is, on this view, in principle possible to construct a completely descriptive concept – i.e. without evaluative force – that picked out the same features of the world.

This account of thick concepts has been criticized by other theorists.\(^{34}\) In their view, which I correspondingly call non-naturalist, the only way to understand a thick concept is to understand the descriptive and evaluative aspects as a whole. The idea is that for a thick concept, the evaluative aspect is profoundly involved in the practice of using it; one cannot understand a thick concept without understanding also its evaluative point.\(^{35}\) Therefore, descriptive terms cannot completely fill in the ‘along the lines’ of a description such as ‘...intended to act in the face of danger to promote a valued end’, the descriptive characterization provided above for courage. Although such descriptions may indeed be

\(^{32}\) Catherine Elgin calls this the “skeleton account” in Elgin 2005, 343.
\(^{33}\) Williams 1985, 141.
\(^{35}\) The rationale for calling an action cruel rather than merely describing it in more neutral terms is to tune in to this evaluative aspect. Cf. McDowell 1981.
crude approximations given to the novice concept monger, providing initial guidelines for grasping the salient features, only a hooking on to the evaluative perspective in which the concept has its point enables a person to apply the thick concept correctly. In the thesis, via investigating a practice involving the risk and safety concepts, I argue for this latter, non-naturalist analysis of thick concepts.

2.2. Decision theory

Decision theory is an interdisciplinary field of inquiry of interest for mathematics, statistics, economics, philosophy and psychology. It covers a broad spectrum of questions involving decision-making, from experimental studies and theories of how we do in fact make decisions (descriptive decision theory) to theories about how we should make them (normative decision theory).\(^{36}\) Even though the foundations were laid in the seventeenth and eighteenth centuries by authors such as Pascal, Bernoulli and Condorcet, modern decision theory dates from the beginning of the twentieth century and gained in influence in the 1950s and onwards.\(^{37}\)

A fundamental concern in decision theory is information on which decisions are made. Classical decision theory divides decision problems into different categories. One category is called decision-making under certainty, referring to cases when we know what the outcome will be for all possible alternatives. Many decisions that we make are – or can at least be approximated as – decisions under certainty. If I have the desire for a mango and contemplate whether to go to my local supermarket to buy one, the decision may not be characterized as a decision under certainty, since the store is sometimes out of mangos. If I, on the other hand, call my friend that works there and get a positive answer that the store is full of them, I know the outcomes of my alternatives: a walk to the store, less money and a mango if I decide to go, and no walk, more money but no mango if I do not.

There are many ways of categorizing situations with less than certainty about the outcome. The point of departure here is the seminal work from 1921 by the economist Frank Knight.\(^{38}\) He made a distinction

---

\(^{36}\) Cf. Resnik 1987 for an introduction to the field, and Gärdenfors and Sahlin 1988 for a comprehensive anthology focusing on Bayesian decision theory.

\(^{37}\) Pascal 1670/1961; Bernoulli 1738/1954; Condorcet 1793/1847.

\(^{38}\) Knight 1921/1957.
between on the one hand “measurable uncertainty” and on the other “something distinctly not of this character”. For the first kind he reserved the term ‘risk’. This kind, he claims, “is so far different from an unmeasurable one that it is not in effect an uncertainty at all”. For the other, “unmeasurable” kind, he reserved the term ‘uncertainty’. The entities referred to as measurable or unmeasurable are the probabilities of the different outcomes.

Similarly, in Duncan Luce and Howard Raiffa’s classical textbook, decision-making under risk is defined as when “each action leads to one set of possible outcomes, each outcome occurring with a known probability” and decision-making under uncertainty as when “either action or both has as its consequence a set of possible specific outcomes, but where the probabilities of these outcomes are completely unknown or not even meaningful”.

Decision-making under certainty, risk and uncertainty remain the three basic categories in classical decision theory. As they have been described by Luce and Raiffa, however, they are not exclusive, since the probability can be partially known. We may, for example, know that the probability that a chess game between Gary Kasparov and Viswanathan Anand will end by a tie or a win for Kasparov when he is holding the white pieces, is, say, 0.6-0.7, but lack more precise knowledge. Then we are not at all completely ignorant of the probabilities, but they are not known with total precision, either. Some textbooks, like Michael Resnik’s Choices, reserve the third category for cases where the probability of the outcomes is unknown or only partly known. I will use the concept decision-making under uncertainty in this latter sense, including partial as well as no knowledge of the probabilities.

---

40 Ibid.
41 Ibid, ch. VII.
42 Luce and Raiffa 1957.
43 How to understand these types of statements is highly controversial in decision theory. Part IV of Gärdenfors and Sahlin 1988 deals with some suggestions for how to conceptualize “unreliable probabilities”.
44 Resnik 1987, 13-14. However, he somewhat confusingly calls this category “decision under ignorance”, an expression that is more commonly used to mark out the case when there is not even partial knowledge of the probability. I will stick to the more common term “uncertainty”.

15
This distinction between decision under risk and decision under uncertainty is fundamental in classical decision theory, where the *probability* referred to is thought to be an objective concept, a property of the world itself. An alternative is to construe probability as a subjective concept. In Bayesian decision theory, probability is conceived of as a measure of the degree of belief that an agent has in a proposition or a state of affairs (such as, say, ‘it will rain tomorrow’). This is combined with a notion of utility into a sophisticated decision system. Frank Ramsey was the first to show that given some rationality assumptions and assumptions on ordering of utilities, it is possible to represent the beliefs of an agent by a unique probability measure. Authors such as Bruno de Finetti, Richard Jeffrey, and Leonard Savage have suggested alternative axiomatizations and developments of Ramsey’s work.

On a Bayesian construal, all (rational) decisions are decisions under risk (known probabilities), since the rational decision-maker always, at least implicitly, assigns a probability value to an outcome. Faced with new information, the agent may change her probability assessment (in accordance with Bayes’ theorem), but she always assigns determinable probabilities to all states of affairs. Critics, internal as well as external to the Bayesian framework, have challenged the plausibility of this view. Daniel Ellsberg, Henry Kyburg, Isaac Levi, Peter Gärdenfors and Nils-Eric Sahlin and others have pointed out that there seems to be a significant difference between some decision situations that should be given the same probability distribution according to classical Bayesianism. The amount of knowledge may vary with the situation at hand, and this seems to be relevant for decision-making. Judging situations such as coin-tossing and the likelihood of picking a red ball from an urn with a known configuration seems very different from judging whether a bridge will hold or what the weather will be like in Rome a month from now. Thus, there seems to be a difference in how certain we can be of the likelihood of an outcome depending on the

---

45 Relative frequencies or logical (“a priori”) probability. E.g. Knight 1921/1957, 214-216, 223-224.
46 Ramsey 1931.
information and situation at hand, i.e. there is an *epistemic uncertainty* that may not be reducible to a unique probability value.

With the inclusion of epistemic uncertainty into the Bayesian framework the gap between classical decision theory and its Bayesian counterpart is considerably narrowed. One common problem is how to specify this epistemic uncertainty, the entity Knight described as ‘unmeasurable’. A significant amount of work has been devoted to this task, and Section 4 of *Essay IV* provides a short overview.

A central tenet in Bayesian decision theory, dominating also in classical decision theory, is the notion of *Maximizing Expected Utility*. The common idea is that each outcome can be assigned a numerical value signifying the goodness of the outcome (the ‘utility’) as well as a probability value, and that the decision-maker should pick the alternative that has the largest sum of the products of the utilities and the probabilities of outcomes. With the inclusion of epistemic uncertainty into the framework, this decision criterion has been questioned and others have been proposed.\(^49\) There is however a fundamental problem with utility ascriptions, namely the strong assumptions that must be in place in order for any criterion even remotely similar to the principle of maximising utility to be meaningful. It is quite easily shown that the utility numbers that are assigned to an outcome must conform to a cardinal scale. This means that we should not only be able to rank an outcome as better than another, but also tell how much better, viz. the relative ‘distance’ between them. This is an assumption that has received a great deal of criticism.\(^50\)

Most situations of risk and safety should be considered as situations under various degrees of uncertainty. The possibility of a harmful event taking place is at the heart of the safety issue. Therefore, it should not come as a surprise that probability and the comparison of utility are important topics in risk and safety analysis.\(^51\)

\(^49\) Cf. Gärdenfors and Sahlin (1982) for an example as well as comparisons with earlier attempts.

\(^50\) Apart from criticism regarding the possibility for one agent of comparing outcomes, there is the question of how to compare utilities among different agents. This is known as the problem of interpersonal comparisons. Cf. Harsanyi (1976), Sen (1970) and Weirich (1984) for some influential views on the subject.

\(^51\) Note that this goes for critical accounts as well as supporting ones: although the degree to which the severity of harm may be construed as (dis-)*utility* is controversial, the decision
3. Looking forward

For practical reasons, there are several issues that I have not been able to pursue in the thesis. As noted earlier, there are two prevailing but largely disparate views of the concepts of risk and safety in risk research; one according to which risks are perceived as objective properties of the world independent of individual beliefs, and another according to which risks are something essentially subjective or socially constructed. A central claim of the thesis has been that that risk and safety are evaluative concepts exhibiting an irreducible normativity. I have argued that this has important consequences for risk research, since our methods of investigation into evaluative concepts differ in important ways from those involving only natural concepts.

Still, I do not think that this puts us in the other, subjectivist or socially constructivist camp: the risk and safety discourse, just as the moral discourse, may be essentially normative without being a mere subjective or social construct. Current developments in moral philosophy have shown that many evaluative concepts bear substantial similarities with basic scientific concepts, bearing ‘marks of objectivity’ that render them more robust than merely subjective opinions. The articles in this thesis have investigated such marks of objectivity, and have taken initial steps towards a realist construal of the concepts of risk and safety as well as the whole class of thick concepts to which they belong. Hence, I have categorized the analysis as ‘non-naturalist’ to mark the affinity to the moral realist view with the same name.

However, in the thesis I stop short of demonstrating realism per se. I have mainly focused on the semantic and epistemological questions involved with understanding and getting knowledge into the thick concepts of risk and safety. This is in line with my preference for what perhaps may be called methodological pragmatism: if we want to gain insight into the metaphysics of an entity, we should start with the functional role of the concepts we use for capturing this entity. Or phrased differently, by starting with the semantic and epistemic function served by the concepts of risk and safety, the metaphysical constraints

theoretical concept of utility is an important point of departure for any account that purports to construct a probabilistic notion of risk and safety.
become visible.\textsuperscript{52} Visible or not, the limitations of the current project have not permitted a full account of these metaphysical constraints. However, I believe that further work will be important for at least two reasons. First, a developed realist view on risk and safety, closely tied to moral realism, may contribute to the conciliatory task of narrowing the gap between natural science accounts of risk and safety and social constructivist accounts, while at the same time bringing in investigative methods from moral theory into the natural science perspective which dominate in risk and safety analysis. Secondly, I believe that such an investigation may be beneficial for several fundamental areas of philosophy as well, notably metaethics and philosophy of science. In particular, the prevailing questions of the status of morality, including the debate between naturalism and non-naturalism, hinges on fundamental issues in philosophy about what reality consists in, how knowledge is gained, and what we mean by our utterances. As argued in the thesis, the well-developed practice of risk and safety research and the debate surrounding it provides an important setting for further investigation into these fundamental questions.

Finally, more should be said about the methodological consequences of the proposed non-naturalism of the concepts of risk and safety. Although I hope that I have been able to show the relevance of the proposed view for risk research, the practical consequences to be drawn are worthy of a more detailed treatment. On a non-naturalist account of risk and safety, such consequences should include structured normative analysis, and tools developed in the field of applied ethics may therefore be applied and developed for usage in risk analysis. A major part of such a task, e.g. developing actual guidelines for risk analysis, is naturally outside the area of philosophy. Still, I believe that there is an important philosophical task in the interdisciplinary work of developing frameworks for such guidelines.

4. Preview of essays I-VI

Essay I. In the first paper, the class of concept to which risk and safety belong – thick concepts – is analyzed. Thick concepts such as courage and cruelty seem to combine descriptive and evaluative features, thus

\textsuperscript{52} Cf Lance and Little (2007).
providing a focal point for several important issues in ethical theory. *Naturalists* analyze thick concepts as the conjunction of a factual description and an evaluation. *Non-naturalists* claim that there are no descriptive boundaries delimiting a thick concept. Traditionally, either of these claims is supported by armchair intuitions about various everyday thick concepts. A more thoroughgoing strategy employed in this paper consists in studying a relevant actual practice, focusing on the competence involved in successful application of the thick concepts in question. I argue that naturalism fails to capture the essential interconnection between the descriptive and evaluative aspects needed for correctly applying the concept, as well as accounting for the difference between explanatory efficacious and merely background conditions. Furthermore, I argue that the two main meta-theoretical worries evoked against non-naturalism – that non-naturalism cannot account for disagreement and that it is not genuinely explanatory – can be met, and consequently that there are no semantic or epistemological trump cards forcing us to believe that naturalism must be true, even when we cannot see how.

**Essay II.** The second paper investigates the most common contemporary critique of the Open Question Argument, the argument utilizing the Kripke/Putnam causal theory of reference. G.E. Moore’s Open Question Argument has been heavily debated ever since it was presented over a hundred years ago. It is argued that an application of the causal theory of reference to moral kinds, rather than defusing the Open Question Argument, actually *underscores* the non-naturalist conclusion. The naturalist critique depends heavily on the analogue between natural kinds and moral kinds. It is argued that the Open Question Argument provides *prima facie* evidence against the claim that moral kinds are natural kinds, and that the causal theory, as interpreted by leading naturalist defenders, does not overturn this evidence. In conclusion, the combination of moral realism and causal regulation theory supports non-naturalism rather than naturalism.

**Essay III.** The third paper applies the previous interpretation of the Open Question Argument to the thick concepts of risk and safety. Within risk research as well as applied risk analysis, risk is most commonly defined as the expected value of the harmful outcome. This conceptualization is a *naturalistic reduction* of risk – the concept of risk is
reduced to natural concepts such as probability and harm. The aim of this paper is to argue against such naturalistic reductions of risk and safety. Three different normative aspects of risk and safety are put forward – epistemic uncertainty, distributive normativity and border normativity. By defending and utilizing a version of Moore’s Open Question Argument, it is argued that these normative aspects cannot be reduced to a natural measure. As a consequence, it is concluded that risk analysis must always contain an openness for the eventuality that there are normative aspects present in a risk situation that are not covered by the operationalizations and models used in risk assessment.

**Essay IV.** The fourth paper of the thesis was written in collaboration with Sven Ove Hansson and Martin Peterson. It provides a conceptual analysis of safety in the context of societal decision-making, focusing on some fundamental distinctions and aspects, and argues for a more complex notion than what is commonly given.

Although safety is a fundamental concept in societal decision-making, it is heavily under-theorized, seldom given any explicit definition or clear characterization in the literature. When it is defined, however, it is often as the antonym of risk – interpreted as a probabilistic notion – an analysis we deem unsatisfactory. The paper explores the distinction between absolute and relative safety, as well as that between objective and subjective safety. Four potential dimensions of safety are discussed, viz. harm, probability, epistemic uncertainty, and control. The first three of these are used in a proposed definition of safety, whereas it is argued that control should not be included in a reasonable definition of safety. It is argued furthermore that, strictly speaking, an objective safety concept is not attainable. Instead, an intersubjective interpretation is proposed that brings us as close as possible to an objective notion. The paper concludes with some formal definitions showing a way of including an interpretation of uncertainty as degrees of confidence.

**Essay V.** The fifth paper explores consequences of epistemic uncertainty for the areas of risk and safety. It is a common opinion in risk research that the public is irrational in its acceptance of risks. Many activities that are claimed by experts to be safe are not deemed to be safe by the public, and vice versa. The paper puts forward a normative critique against a common argument, viz. that the public should follow the experts’ advice in recommending an activity whenever the experts
have the best knowledge of the risk involved. Even after making plausible limitations to exclude ‘external’ considerations, the claim remains incorrect. The importance of safety in risk acceptance, together with the phenomenon of epistemic uncertainty, highlights the vital concern, not whether the expert knowledge of the risk is the best one available, but whether that knowledge is good enough. This introduces an ‘internal’, yet extra-scientific, value component, invalidating the claim. Furthermore, I show that the scope of the objection covers the entire field of risk research, risk assessment as well as risk management.

**Essay VI.** The final paper, written in collaboration with Sven Ove Hansson, provides a systematized account of safety engineering practices that clarifies their relation to the goal of safety engineering, namely to increase safety. A list of 24 principles referred to in the literature of safety engineering is provided, divided into four major categories:

1) **Inherently safe design.** Minimizing inherent dangers in the process as far as possible. Potential hazards are excluded rather than just enclosed or otherwise coped with.

2) **Safety reserves.** Constructions should be strong enough to resist loads and disturbances exceeding those that are intended. A common method is to employ numerical safety factors.

3) **Safe fail.** The system should fail ‘safely’; either the internal components may fail without the system as a whole failing, or the system fails without causing harm.

4) **Procedural safeguards.** Procedures and control mechanisms for enhancing safety, ranging from general safety standards and quality assurance to training and behavior control of the staff.

It is argued that important aspects of these methods can be better understood with the help of the distinction between risk and uncertainty, in addition to the common distinction between risk and probability. On a narrow risk reduction interpretation of safety (understanding risk as the combination of probability and severity of harm) we cannot fully account for these principles. This is not due to deficiencies of the principles, however, but to a shortcoming in the capability of such a theoretical framework to capture the concept of safety.
References


