ABSTRACTIONS AND IDEALISATIONS IN EPISTEMIC LOGIC

MSc Thesis *(Afstudeerscriptie)*

written by

Ottilia Kasbergen
(born October 2nd, 1989 in Loosdrecht, The Netherlands)

under the supervision of **Prof Dr Sonja Smets** and **Prof Dr Martin Stokhof**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
June 14th, 2017	Prof Dr Michiel van Lambalgen
	Dr Floris Roelofsen
	Prof Dr Frank Veltman

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

In this thesis we investigate to what extent the many abstractions and idealisations that epistemic logical models make constitute a problem towards the practice of epistemic logic. To this end we look at examples of abstractions and idealisations in epistemic logic, in particular focusing on the problem of logical omniscience. We find that the acceptability of different idealisations depends on the research context of the model in question. However, in epistemic logical literature we rarely find discussions of aims, success criteria and intended applications. Therefore, we asked ten epistemic logicians to share their views on the practice of epistemic logic. As their views diverged greatly, we propose a categorisation of the field of epistemic logic into four research agendas. These agendas will function as a starting point for a discussion of the advantages and disadvantages of idealisations in epistemic logic. As idealisations are inherent to modelling, and as such not an issue to epistemic logic only, we study modelling and idealisations in the sciences more broadly as well. We find that usage of the terms 'abstraction' and 'idealisation' is not standardised, therefore we redefine these concepts for a discussion of the practice of epistemic logic. We find that it is of importance that epistemic logicians are explicit about both their aims or intended applications, as well as about the idealisations that their models make. Being explicit about the research context may enhance an adequate conception of the extent of use of the epistemic logical models, and may enhance interdisciplinary research.

Acknowledgements

As the Māori say: "*Ehara taku toa, he takitahi, he toa takitini.*" My success is not my own, it is the success of a community. This means that our achievements do not belong to us alone, but also to those around us. Our friends, family, colleagues and the whole community make us what we are. We stand on the shoulders of giants, as Newton rightly remarked. However, those giants are not only geniuses of science, but also the everyday people around us. They may lift us up, show us directions and possibilities, guide and inspire us. I am grateful for all the people that I've met, for all the beautiful people that surround me both in person and in spirit.

As I am in nature are truly grateful person, and because I don't think I will publish anything soon again and therefore may not have another opportunity to give acknowledgements soon again, I will give a lengthy list of the people and experiences that I am grateful for. It is my favourite chapter of this thesis.

I would like to thank my wise supervisors Sonja Smets and Martin Stokhof, who stayed with me, even though I had to take a break from my project due to personal reasons and begin again a year later. I'm also very grateful to all my dear friends that have helped me in the process of writing this thesis: Hanna, for your extensive help throughout the whole process, always there to discuss this topic and my writings with me; Iris and Ronald, for proofreading my thesis, discussing the content with me, and helping me with LaTeX-figures; Anggha, for helping me with my references without even knowing me before; Brendan, Brian, Iris and Thomas, for all your hours spent on improving my English phrasing and making this thesis readable.

I would also like to express my gratitude to the thesis defence committee chair Floris Roelofsen, and committee members Michiel van Lambalgen and Frank Veltman, for reading this thesis and for playing your role in my graduation. Additionally, I would like to thank Floris for the many good conversations we had during my Master's degree and for your academic guidance.

I thank Johan van Benthem, whom I discussed the subject matter with already early in this process, and I thank all the other researchers that I interviewed for this thesis project: Alexandru Baltag, Marta Bílková, Joseph Halpern, Vincent Hendricks, Nina Gierasimczuk, Hannes Leitgeb, Benedikt Löwe, Rohit Parikh and Jan-Willem Romeijn. Besides truly inspiring, I feel that conducting the interviews has also been a lot of fun. I am grateful that you all took the time for our conversation, and shared your views with me.

I would like to thank my New Zealand academic mentor Willem Labuschagne, who inspired me to look at logic in a whole new way. I am also grateful to Tanja Kassenaar of the ILLC office, who makes life for the Master of Logic students so much better. I hope you recover soon. Furthermore, I am grateful to all those people that I met during my time at the Institute of Logic, Language and Computation at the University of Amsterdam, and during my visits to the University of Otago and the University of Auckland.

Besides all those people who helped shape me academically, I also wish to give thanks to those who surround me personally. First, I would like to thank my family: my mother and father, sisters and brother, and those many uncles and aunts, cousins, nieces and nephews, who are supportive and to whom who I always enjoy talking. I'm very grateful for having a large extended family as well: Sanne, Simone and Gerben, Ingeborg, Lana, Kasper, Joachim, Tim, Roel and the kids; Judith, Anke, Sien, Sjoerd, and all the other Stoepers; Houssam, Jorim, Vivian, and all the other DenkTankers; Ana Lucia, Eileen, Nigel, Adu, Aldo, Suzanne, Kees and all the others of the MoL-gang; Ruben, Kim, and the rest of the Betabruggers, my colleagues of OverBruggen in Amsterdam; Kim and all the others of our big and beautiful rabbit family.

I am grateful to all the people that I met during my travels and that found themselves a place in my heart (I had to expand its size). In particular, I am grateful for getting to know (in chronological order) Alexander, Felix and his fiancee Francesca, the Fischer family, Beti and her family (aunt Jožica), Tina and her friends, the homeless person in Piran who re-taught me to see and enjoy what there is (his favourite prime-time show being the sunset over the bay), Astrid and Oliver, Amina's mother, Alen, my student family of Boszarkany, Kolja, Anke und Andreas, Carlos, Justyna, Olia, Kate and Sasha, Alina, Arthur, my Sasha, Sveta and her friends, the nuns of Achairsky monastery, Natalya's family and friends, Emma, William, Baggi, Naka, Ghanaa and her family and the other Mongolian nomads that showed me so much hospitality, Babushka (Sashina), the Saudade family, the Taia whānau, Flor, my Dunedin flatmates and my international student friends; Jess, Carl, Moss, Max and Kotty, Brendan and his family, Judy, my bone carving teacher Pav and my jade carving teacher Kendall, auntie Joan, Papa G, the Kingsway school kids, Te Rarawa, Te Rereahu, and all those still to come.
They taught and teach me so much about life, love, kindness and hospitality.

It is because of them that I managed to write this thesis, and probably also (partially) because of them that it took me so long.

As for the rest, I'm grateful for all those things that you often forget you have when you grow up in a wealthy and peaceful country like the Netherlands: peace, safety, human rights, enough food, good health care, and so on. Last but not least: I'm grateful for having music in my life. And of course for being able to write such an extensive list of people and things I'm grateful for and meaning every word of it. (meta-grateful)

Contents

Chapter 1

Introduction

1.1 Foreword

When I presented some preliminary results of my thesis project, someone asked me why I would not just go back to doing formal work, instead of procrastinating with these metastudies of epistemic logic. Spending so many hours on this thesis project, it goes without saying that I do not agree with a metastudy of my favourite field of logic being called 'procrastination'. Although I love solving intriguing puzzles in a more formal way, I do not think that philosophising about the way in which we do that is less valuable or less important. Besides doing your work well, wouldn't it be a vital part of research to also think about why you do the things you do and why you do them in the way you do them? Of course, researchers think about these things, but I noticed that much of this subject matter remains implicit, and that the discussion that there is between researchers is often not written down for others to see. I find it interesting to look at how researchers within the different fields of epistemic logic look at their practice, and fascinating to uncover some of the similarities and sometimes even outright contradictions in their views. I think that we can learn much more from each other when we understand thoroughly where the other comes from and in what way they use the words they speak.

1.2 Philosophy of epistemic logic

When I was doing my Bachelor's in Mathematics at the University of Amsterdam, Johan van Benthem introduced me to modal logic, and consequently to epistemic logic. I was immediately intrigued by these beautiful systems, and a year later I wrote my thesis on the surprise examination, supervised by Alexandru Baltag. I immensely enjoyed working with these epistemic logical systems (so much even that I entered the Master of Logic), but I always kept wondering: What do these systems say about reality? They make such unrealistic assumptions! Sometimes I felt like they were 'just' cool mathematics with nice pick and choose examples that relate them to concepts we know, providing them with a reason for existence.

In 1960 Lewis stated:

> "It is so easy to get impressive 'results' by replacing the vaguer concepts which convey real meaning by virtue of common usage by pseudo precise concepts which are manipulable by "exact" methods – the trouble being that nobody any longer knows whether anything actual or of practical import is being discusseded." (Lewis 1960, cited in: EL5Q[1], 234)

[1]We use the abbreviation 'EL5Q' to refer to the book *'Epistemic Logic: 5 Questions'*, that can be found in the bibliography via (Demey et al. 2011). The book consists of interviews with 25 epistemic logicians. We considered

Lewis was talking here about philosophy more generally, but he captures my confusion over the idealisations of epistemic logic well. What is the practical importance of epistemic logic, and can we say something about anything actual? To what extent do the idealisations of our models constitute a problem towards the practice of epistemic logic? Exactly that is what I always kept wanting to study. So I did. The result is in front of you. It is a work of philosophy of epistemic logic.

1.3 Chapters

This thesis is set up as follows: chapter 2 introduces epistemology and epistemic logic shortly, after which we investigate in which ways epistemic logic deviates from the reality of human knowledge. This also contains a case study of the problem of logical omniscience. We find that we need to know more about the aims and research context of epistemic logic. In chapter 3, therefore, we have a look at the views that epistemic logicians expressed in interviews[2], concerning the practice of epistemic logic: its aims, success criteria, methodology and why or why not idealisations may constitute a problem. In order to get a grip on the diversity of the views of epistemic logicians concerning these matters, we introduce a categorisation of the field of epistemic logic into four research agendas. In chapter 4 we study the topic of modelling in the sciences more broadly, which teaches us some valuable lessons. As the usage of the terms 'abstraction' and 'idealisation' is not standardised, we redefine these concepts in a way that suits the practice of epistemic logic. In chapter 5 all the lessons from the previous three chapters are combined in a discussion, followed by the conclusion and suggestions for further work. Although chapters 2, 3 and 4 focus mainly on describing and interpreting what has been written and said by epistemic logicians and philosophers of science, chapter 5 shows more opinion and some suggestions that emanate from this study.

The text of a thesis is linear, the process of research for this thesis of course not. Therefore, the material was reordered as to get a comprehensible line of argument. In order to not miss out on the valuable connections between chapters 2, 3 and 4 while reading them, footnotes have been added to uncover such connections.

1.4 Methodology

For this thesis material from three sources was used: literature of epistemic logic, interviews conducted with epistemic logicians, and literature of the philosophy of science. Whereas the literature studies are quite straightforward, the process, methods and choices regarding the interviews are introduced more elaborately.

1.4.1 Motivation

As became clear from studying the literature of epistemic logic, discussions about aims and considerations concerning goals, success criteria and methodology of epistemic logic are very rarely written down. However, knowing the research context is important for assessing whether some idealisation or abstraction constitutes a problem towards the practice of epistemic logic.[3] Therefore, there is a need for additional information. To fill this gap researchers working in and around the field of epistemic logic have been asked to share their views in short interviews.

---

an abbreviation necessary, because the book is frequently referred to, and adding '(Demey et al. 2011)' every time after some other researchers' name would be confusing.

[2] The next section further elaborates on the chosen methodology.

[3] This will frequently be argued in the following chapters.

1.4.2 Interview design and process

Number of interviewees During this thesis project ten researchers have been asked to participate in an interview. All of them agreed to do so. Ten is considered here to be the right number for current purposes: it gives a diverse enough view, while still manageable in the alloted time for a thesis project.

Selection of researchers The 10 researchers were selected because they approach epistemic logic differently in their work: doing mainly mathematical work or mainly philosophical, working in the direction of cognitive science, social psychology or computer science; working mainly with quantitative methods or even working outside epistemic logic (but with knowledge of the field). During the selection process attention has also been paid to affiliation and gender. The researchers are affiliated with the following universities: University of Amsterdam, Stanford University, Tsinghua University, Charles University in Prague, Cornell University, University of Copenhagen, Danish Technical University, Ludwig-Maximilians-Universität München, Universität Hamburg, City University of New York, and University of Groningen. Among the 10 researchers are two women and eight men.

The interviewees are Alexandru Baltag, Johan van Benthem, Marta Bílková, Joseph Halpern, Vincent Hendricks, Nina Gierasimczuk, Hannes Leitgeb, Benedikt Löwe, Rohit Parikh and Jan-Willem Romeijn.

Length of the interviews The interviewees were asked to participate in a 25-minute interview. Some of them chose to spend a little more time on the topic.

Interview dates The interview were held between February 3rd and March 6th, 2017.

Interview format The interviews followed a semi-structured format. A set of questions has served as a starting point for the conversation. Where needed, follow-up questions were asked to elucidate answers or reach more depth.

Questions The questions of the interview were sent to the interviewees beforehand (attached to the email with the invitation), so that they had the opportunity to ponder on them before participating in the interview. The list of questions can be found in Appendix C. They are open-ended, as a range of possible answers was to be expected. The questions had the function of exploring the subject material in considerable depth on the basis of the experiences of the interviewees.

Process When the questions were carefully selected for usefulness and efficiency, a tryout interview was held with PhD student Hanna van Lee. The set of questions was found appropriate both in topics and in length, and based on this tryout the sem-istructured format was chosen. In the following month, two or three interviews were conducted each week. The interviews were conducted either by computer via the free software package Skype (when the interviewee lived far away) or in person (when the interviewee was in Amsterdam). All interviews were recorded and summarised. A few days after conducting the interview, this summary (two or three pages long, containing both summarised viewpoints and citations) was sent to the interviewee, upon which they gave their permission to use the material, in some cases accompanied by some clarifications or adjustments. In this way the interview material has been formed into written statements that the interviewees agree with.

1.4.3 Justification

As the interviews yield written statements that the interviewees approved, the interviews provide us with useful material for this thesis. The goal was to get more information about researchers' opinions on the practice of epistemic logic, as this is hardly ever discussed in written text. Getting this extra information is exactly what was achieved by conducting the interviews. Conducting interviews among researchers was found to be the most appropriate research method for the current goal.

1.4.4 Generalisation

The material from the interviews is not meant to engender claims about how all or most of epistemic logicians think about this subject matter. The interviews are merely meant to be illustrative of how one *may* think about the subject matter, and how the interviewees *do think* about it. In this thesis the information that the interviews yielded has been solely used instrumentally for the purpose of saying something meaningful about abstractions and idealisations in epistemic logic.

Chapter 2

Epistemic Logic and its Representational Imperfections

Epistemic logic is a way of doing epistemology. The word 'epistemology' is a combination of the Greek word 'episteme', which means 'knowledge', and the familiar suffix '-ology', signifying 'study of'. Epistemology is sometimes also referred to as 'theory of knowledge'.

Epistemologists are concerned with questions of the following kind: What is knowledge? How does knowledge differ from belief? How can knowledge be justified? Is knowledge always truthful? Is knowledge monotonic, and do we, once acquired, never lose it? What are necessary and sufficient conditions for achieving knowledge?

Answers to questions related to the notion of knowledge can be sought for in different ways. Roughly, we can distinguish between 'mainstream' and 'formal' approaches in epistemology. Although there seems to be a great divide between mainstream epistemology and formal epistemology, their starting point is the same. The main difference between the two strands of epistemology is a difference in their modus operandi, described by Vincent Hendricks as follows:

> "'Mainstream' epistemology ... seeks necessary and sufficient conditions for the possession of knowledge using largely common-sense considerations and folksy examples and counterexamples." (Hendricks 2005, 4)

> "'Formal' approaches to epistemology ... either proceed axiomatically or concentrate on learning and knowledge acquisition using toolboxes from logic and computability theory." (Hendricks 2005, 5)

Both Hendricks and Johan van Benthem have argued that the apparent divide between mainstream and formal epistemology is not as big as it seems, and that the approaches throughout history often meet. Both are of the opinion that such connections, or 'mergers'[1], are very fruitful, and that mainstream and formal epistemology each have advantages that can bring research in both fields further.[2]

As Wolfgang Lenzen (1978, 17) writes: "most principles of epistemic logic can be evaluated only in the light of epistemological considerations," and that is exactly what we purport to do in this chapter. There are many debatable topics and considerations in epistemology that have bearing on both mainstream and formal approaches. Before continuing to discuss formal epistemology

[1] Term used by Hendricks in his book *'Mainstream and Formal Epistemology'* (Hendricks 2005).

[2] See for a more elaborate discussion of this Hendricks 2005, Hendricks and Symons 2006 or Van Benthem 2006.

more closely, and, in particular, epistemic logic, we have a look at a few of those topics of debate. However, we should remain aware of the fact that the field of epistemology is much larger than this[3], and that we are just looking at some selected snippets of epistemology that will prove themselves useful later on.

2.1 A short introduction to epistemology

The word 'know' is very frequently used in the English language. To be precise, 'know' ranks 59^{th} of all word forms, and it even gets the 8^{th} place among English verbs, according to an analysis of the Oxford English Corpus.[4] However, not all uses of the word 'know' interest us momentarily. In ordinary life one can often refer to knowledge in a very loose manner –"He's gonna win! I just know it!"–, but this is not the form of knowing that epistemologists usually investigate. Rather they are interested in a more strict notion of knowledge, in which knowledge is not-metaphorically attributed to someone. Even if we consider only knowledge in this strict sense, there are still a lot of different possibilities. Compare the following sentences:

(1) "Did you know *that it rains outside*?"

(2) "Alina knows *the poem 'Winter morning' by Pushkin* by heart."

(3) "I know *aunty Joan* very well."

(4) "Of course they know *how to tie their shoelaces*."

'To know' is a transitive verb, it has a subject and an object. The epistemic subject, otherwise referred to as 'agent', is 'that what knows', commonly human beings, but also computers, God(s), animals, et cetera. The epistemic object, then, is 'that what is known', which can be any state of affairs (as in sentence 1), particular thing (as in sentence 2) or person (as in sentence 3), or procedure[5] (as in sentence 4). These types of knowing can respectively be referred to as 'knowing that', 'knowing what' or 'knowing who', and 'knowing how'. Most attention in epistemic logic goes out to the first type, 'knowing that', hence this will be the focus of this thesis.

2.1.1 Definitions of knowledge

Now that we have roughly outlined what kind of knowledge interests us, we can raise the question: what exactly *is* this kind of knowledge? As history of philosophy teaches us, this question is not so easy to answer. In fact, even today epistemologists keep having disagreements about how to define knowledge.

In Classical Greece, Plato mooted a good candidate for a definition of knowledge that lasted for many centuries: justified true belief.[6] Just 'true belief' was not enough, because if someone happens to be accidentally right in his belief, we would not want to call it knowledge. However, Gettier did away with this definition in his famous 3-paged essay in 1963 that included two simple

[3] We stay largely silent about issues like the internalism-externalism debate, contextualism, sources of knowledge, and many other fields of epistemology, e.g., social epistemology, religious epistemology, and feminist epistemology.

[4] Results of the analysis published on Oxford Online. The Oxford English Corpus encompasses only written material, but for this particular point it is instructive of general usage.

[5] In fact, we can further distinguish between procedural knowledge and performatory knowledge: we might be able to catch a ball without being able to exactly say how it is we do that. In this case we have performatory knowledge, not procedural knowledge. (Rescher 2005, p5)

[6] More specifically, Plato argued that true belief requires some 'logos' in order to count as knowledge. However, this is traditionally translated or referred to as justified true belief.

counterexamples. Not only does the true belief need to be justified, it needs to be justified in the right way. (Gettier 1963) What precisely this missing ingredient in the definition of knowledge is has been debated ever since. A year earlier, Jaakko Hintikka (1962), in his famous book *'Knowledge and Belief'*, coined a whole new way of defining knowledge: truth throughout all possible ways the world could have been. Other influential suggestions have been done by, for instance, Dretske (1981), interpreting knowledge in terms of information, and Nozick (1981), with his truth-tracker account of knowledge.

All these definitions or interpretations shine a different light on the concept of knowledge. What they have in common, as Van Benthem (2006) notes, is that they all define knowledge in terms of some further basic notions, such as belief, truth, justification, information and counterfactuals. Consequently, a study of epistemology also involves investigation of these notions.

One of the features of knowledge that is often mentioned is its 'stability', or, 'robustness'. As Van Benthem (2006, 56) articulates it: "Knowledge is not attained lightly, and the way it was acquired guarantees a certain stability through changing situations." This robustness has historically and recently received much attention from epistemologists, and it leads us into the discussion of fallibility versus infallibility. How strong does the epistemic position of an agent have to be in order to be attributed knowledge? Proponents of infallibilism contend that the agent's epistemic position has to be so strong that she could not go wrong in knowing the proposition in question. Fallibilists, however, argue that this condition is too strong, consequently the agent can be mistaken. Mattey (2015) has argued that nowadays most epistemologists are fallibilists, in consonance with real-life attributions of knowledge.

When trying to find the best definition for knowledge, one may ask what is the right way around: either to start with the recognition that people have knowledge, and then try to figure out what its characteristics are[7], or to start with a prior conception of what knowledge is, and then figure out if people have knowledge or not. The first method has been called "particularism", and the second "methodism."[8]

A counterargument against particularism is that it is impossible to identify items of knowledge among people if we do not have a prior conception of what knowledge means already. However, the particularist may reply, countering methodism, that it is also impossible to have a conception of knowledge if we have not yet determined what kind of items count as knowledge. Clearly there seems to be some circularity in both methods, and therefore it is hard to resolve this issue with arguments that are not ideological in nature. The particularist may resort to observations of when people in practice attribute knowledge to someone, but this varies from person to person and from situation to situation. The methodist may resort to ways of defining knowledge that philosophers have coined, but –as we have already seen– also among those there is a great variability.[9]

2.1.2 Descriptive versus normative

In investigating knowledge, we may distinguish between questions of fact and questions of value. The former require answers in terms of what *is*, whereas the latter require answers in terms of what *should be* or *ought to be*. The first approach is called the 'descriptive' approach, whereas the latter sets up norms for knowledge and is called 'normative'. In epistemology one rarely encounters

[7]Here we can further distinguish between the approach of defining knowledge in terms of belief (or another epistemic concept, like information) plus some extra condition, and the 'knowledge first'-approach, that considers knowledge to be a concept that cannot be defined in terms of another epistemic concept. Timothy Williamson advocated the latter approach, see (Williamson 2002) and (Williamson 2013).

[8]In Chisholm 1973, cited by Mattey (2015).

[9]Discussion following the line of Mattey 2015.

purely descriptive or purely normative approaches. The reason for this is that for describing the situation as it is, one needs to define certain concepts (e.g., what counts as knowledge), and, likewise, it is hard to set up norms for knowledge without describing knowledge and situations in which knowledge occurs. Most work in epistemology is partly descriptive and partly normative, only the extent to which an approach should be descriptive or normative is debated.

2.1.3 Skepticism

Socr.	*Is it possible, then, to reach truth when one cannot reach existence?*
Theaet.	*It is impossible.*
Socr.	*But if a man cannot reach the truth of a thing, can he possibly know this thing?*
Theaet.	*No, Socrates, how could he?* (Extract from Plato's *'Theaetetus'*, 186c-d)

Virtually all accounts of knowledge are vulnerable to scepticism. The sceptic can argue that acquiring knowledge is impossible, because of the permanent possibility of error. We may never know for sure if we are not mistaken in our (self-)attribution of knowledge: after all we may be dreaming, or a brain in a vat in some labatory. We can never really exclude the possibility that we are in completely different circumstances than we think we are.

Furthermore, we can always be wrong in our justification of knowledge of some proposition. Justification can take place in (at least) three ways: we can deduce the proposition, we can observe the state of affairs that is described in the proposition, or the proposition can have been communicated to us by someone that we trust (to be infallible). The sceptic can defy us in all three cases: he can say that we cannot be sure that our reasoning abilities are sufficient (after all, people often make reasoning mistakes), that our sense perception is not trustworthy (after all, we often are deceived by our senses), and that the person who communicated the proposition to us may have been lying or been mistaken (after all, people often lie or tell untruths).

Of course, the impact of the sceptical argument depends on our definition of knowledge, in particular on how strong we conceive knowledge to be. For infallibilists the sceptical argument is quite disastrous, but if we accept that we might not always be right in our attribution of knowledge the argument has less force.

2.2 Advantages of formalisation

"[P]erhaps not just the Book of Nature, but also the Book of Ideas is written in mathematical language." (Van Benthem 2007, 67)

Before we go on to discuss epistemic logic in greater detail, we should ask ourselves why we would want to use logic to investigate epistemological questions.[10] Or rather, why would we want to make use of formal tools in order to get answers to this sort of questions? In his recent dissertation, Dominik Klein offers us a great starting point to discuss the advantages that formal tools, or logic in particular, may bring. (Klein 2015) In formalised sciences, instead of studying the 'target system', the real-world phenomena, object or situation that we want to investigate, we study a 'formal system'. Klein claims that this can benefit us in three main ways: clarification,

[10] In order to answer this question, one has to know what logic *is*. A straightforward definition of logic seems hard to give, and even among logicians there seems to be little agreement on this. Deteriorating the problem is that one can define logic narrowly as the study of right reasoning, but most researchers take the topic to be much broader defined. Van Benthem (2007, 94) remarks that "'logic' is better described as a certain rigourous style of working than any very precisely demarcated subject matter." This seems to be a great way of putting it. We do not desire to give a (more detailed) definition of logic, and will (realistically) assume that the readers of this thesis have a good enough grasp of what logic is.

verification and exploration. We will discuss the benefits of using logic (in epistemology) along these lines.[11]

Clarification. The most mentioned and noticeable advantage that using logic has for epistemological questions, is that it provides notation with the help of which a problem can be stated very precisely. The problem in question is thereby explicated, and ambiguity or vagueness reduced. This clarifying role of logic is articulated by Van Benthem (2007, 67) as follows:

> "Logic is not applied to philosophical problems the way an engineer might apply some technique for computing stress in a bridge. Its role seems much more indirect. Logic offers a technical language with relatively precise meanings as an enhancement of philosophical discourse, and an aid to precise communication."

However, not only are the ambiguities reduced by a precise statement of the subject matter, the precision of the technical language may also constitute a revelation of ambiguities or gaps in our philosophical discussion of the problem, so that the investigation of the problem is enhanced. Furthermore, as Klein suggests, using logic may throw new light at a problem by drawing attention to structural patterns, for example by identifying learning possibilities of the agent. A great advantage of the use of logic is that some selected aspects of an epistemological question can be studied in isolation in greater detail. (Klein 2015, 175)

Verification. Because of the precise formulation that a logical framework asks for, underlying assumptions and preconditions will be made explicit, and are thereby up for evaluation. Furthermore, with the help of logic one can identify necessary or sufficient assumptions needed for deriving a particular conclusion. Klein contends that, because philosophy is made more precise with the help of logic, inferences can be assessed with respect to their validity and it can be verified if certain expected relationships indeed exist. (Klein 2015, 176)

Exploration. The frameworks that are used in epistemic logic are equipped with mathematical and logical tools[12] that enable us to explore the problem further, and thereby provide us with additional insights. This can be compared to experiments in empirical science, but then these experiments are conducted in some logical framework. However, there are considerable differences between experiments in empirical science and demonstration in a logical system, as will be discussed in section 5.1. Van Benthem (2006, 71)) mentions that logic can also aid in constructing new philosophical conceptual frameworks. Moreover, using logic may reveal analogies between different notions or problems that one was still unaware of. (Van Benthem 2007, 67)

However, there are not only advantages to using formal methods. In section 2.4 and 2.5 we will discuss some possible shortcomings of epistemic logic, a form of formal epistemology. Before we can have a look at these idealisations, we will introduce epistemic logic.

2.3 An introduction to epistemic logic

Epistemic logic is a form of formal epistemology, the study of knowledge and reasoning using formal tools, that is, using mathematical, computational or logical techniques. One could, for example, study knowledge or belief in terms of probability, or use game theory to investigate communication and information flow. Another booming field of formal epistemology is the one that intersects with computer science.

[11] Cf. section 3.3.4. In this section an overview is given of what the interviewed researchers consider to be the advantages of using epistemic logic.

[12] Cf. the discussion of 'internal dynamics' in section 4.1.4.

Subdisciplines of formal epistemology

Formal epistemology is concerned with questions similar to those investigated in mainstream epistemology, but often the exact terminology is different. For example, the mainstream epistemologist may be speaking about at a defeasibility condition, the epistemic logician considers non-monotonic logics and default rules; when the mainstream epistemologist talks about getting to the truth eventually, the computational epistemologist thinks about successful convergence and solvability, et cetera. (Hendricks 2005, 6) However, the borders between the subdisciplines of (formal) epistemology are not as rigid as they may seem. It is important to be aware of the connections between different fields, because the study of knowledge transcends borders of disciplines and methods. Van Benthem warns us for the trap of fixed subdisciplines and system imprisonment (Van Benthem 2007, 66): rather than getting stuck on borders between various philosophical subdisciplines and between various formal systems, we should think in terms of themes. If we do so, we are less likely to get stuck on limitations of a particular discipline, hence science will progress more quickly. However, having that said, due to limits in time and pages, from now on we will focus on epistemic logic only.

2.3.1 A short history of epistemic logic

Although many interesting ideas had already been put forward in connection to a systematic treatment of knowledge in some axiomatic-deductive system, in particular Carnap's *'Meaning and Necessity'*, von Wright's *'An essay on modal logic'*, and work done by Prior and Rescher, generally Hintikka's *'Knowledge and Belief'* is viewed upon as the main work leading to epistemic logic[13] as we have it today.

Hintikka (1962) suggested that we can take Kripke's modal logical framework[14] –that had recently been developed–, and give it an epistemic interpretation. Knowledge is now seen as truth throughout all possible ways the world might have been, by Hintikka referred to as 'scenarios'. Nowadays these scenarios are commonly referred to as 'possible worlds', but Hintikka himself has recently been critical of this usage:

> "In order to speak of what a certain person a knows and does not know, we have to assume a class (space) of possibilities. These possibilities will be called scenarios. Philosophers typically call them possible worlds. This usage is a symptom of intellectual megalomania." (Hintikka 2003, 19)

Since Hintikka's seminal work, the field of epistemic logic has flourished and many interesting contributions have been made. Throughout the decades, formal treatment of knowledge has become more and more realistic by adaptions and additions to the framework. Because of this, we can now formally investigate dynamics, multi-agent cases, different degrees of strength of knowledge and belief, dishonesty, various trust attitudes, relevant alternatives, et cetera.

The logical landscape is broad. Throughout the years, standard epistemic logic has been extended and adapted in order to make it more realistic or fitting for particular purposes. We will turn to a discussion of such frameworks in section 2.5 on the problem of logical omniscience, but first we will introduce and discuss standard epistemic logic. In the rest of this thesis we will assume that the reader has sufficient background in modal logics, and in particular, epistemic logic. For an introduction to modal logics we refer to Appendix A, for an introduction of the different epistemic logical frameworks discussed in 2.5 we refer to Appendix B.

[13]The term 'epistemic logic' is often used for the logic of knowledge *and* belief. More correct would be to refer to the logic of belief as 'doxastic logic'. In this thesis we mainly concentrate on the logic of knowledge.

[14]See Appendix A.

2.3.2 An epistemic interpretation of modal logic

In epistemic logic, standard modal logical systems as discussed in the previous section are interpreted epistemically. This has the great advantage that all technical results that are achieved in the standard systems can be immediately transferred to epistemic logic. Syntactically, the modal box operator '$\Box$' is replaced by the epistemic operator 'K'. Semantically, instead of interpreting the operator in terms of necessity, it is now interpreted in terms of knowledge. '$K_i\varphi$' for some arbitrary individual i and proposition φ, now signifies 'agent i knows that φ'. Thus, epistemic formulae like this express the epistemic attitude of some agent towards certain propositions.

Knowledge is defined, as put forth by Hintikka in 1962, as a kind of 'epistemic necessity', as truth among all accessible possible worlds[15]:

Definition 2.1. *(Knowledge). For each possible worlds model* $M = (W, R_i, \| \cdot \|)$*, with* W *a non-empty set of possible worlds,* R_i *the accessibility relation between the worlds of* W *for each agent* i *in* G *a group of agents and valution map* $\| \cdot \|$*, and each world* $w \in W$*:*
$M, w \models K_i\varphi \Leftrightarrow$ *for all* w' *with* wR_iw'*:* $M, w' \models \varphi$

The accessibility relation, as defined in Appendix A, now gives the agent's epistemic alternatives. So wR_iv in this case means that in world w, agent i considers world v to be a possible scenario of the way the world is. Those worlds that are incompatible with the agent's epistemic position, that is, the worlds that cannot be the actual world according to what the agent knows, are excluded. As before, we consider a non-empty set W of possible scenarios. The worlds that the agent has access to from a certain viewpoint are for her indistinguishable from the actual world. That is the reason why the accessibility relation, or epistemic-alternatives relation, sometimes also goes by the name of 'indistinguishability relation'.

As mentioned, the basic epistemic language consists of the propositional logical language plus some epistemic operator 'K'. However, in the epistemic interpretation of modal logic, sometimes also additional epistemic operators are considered. Examples of these are common knowledge 'C_G', distributed knowledge 'D_G', everybody knows 'E', et cetera. Possible extensions of the basic epistemic logical language will be discussed in sections 2.4 and 2.5.

Table 2.1 Epistemic axioms (Based on Garson 2016)

Name	**Axiom**	**R is...**
(K)	$K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$	no specification
(T)	$K_i\varphi \rightarrow \varphi$	reflexive
(4)	$K_i\varphi \rightarrow K_iK_i\psi$	symmetric
(5)	$\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$	euclidean
(2.)	$\neg K_i\neg K_i\varphi \rightarrow K_i\neg K_i\neg\varphi$	convergent
(3.)	$K_i(K_i\varphi \rightarrow K_i\psi) \vee K_i(K_i\psi \rightarrow K_i\varphi)$	weakly connected[16]
(4.)	$\varphi \rightarrow (\neg K_i\neg K_i\varphi \rightarrow K_i\varphi)$	remotely symmetrical[17]

As in Appendix A, different systems result from different sets of axioms. Table 2.1 illustrates the epistemic axioms that are commonly discussed in epistemic logic. We will discuss each epistemic axiom and the philosophical discussions around them separately. All epistemic operators in the axioms are indexed with i for some agent i for the practical reason that it makes it easier for

[15] Recall Hintikka's criticism of the terminology.

[16] See '*Many-dimensional modal logics: theory and applications*' for a discussion of this. (Kurucz et al. 2003)

[17] That is, as suggested by Georgacarakos, the accessibility relation of **S4.4** can be correct arrived by the properties of being convergent and remotely symmetrical. (Georgacarakos 1976)

us to discuss the axioms in an intuitive fashion. However, the epistemic operator would in the single-agent case commonly go without index, being read as ‘it is known that’, and in multi-agent cases be indexed with either a particular index i, referring to one agent, or a group index ‘G’, referring to a group of agents.

Epistemic logic has the means to go beyond just factual knowledge. It also enables formally investigating knowledge about knowledge, which is especially interesting in multi-agent cases. In multi-agent epistemic logic we can express that, for example, Alice knows that Bob knows, that Alice knows that there is a dinner party tonight:[18] $K_a K_b K_a \varphi$.

However the merits of epistemic logic, there are some interpretation problems with respect to natural language sentences. Take a common utterance like ‘Alice doesn’t know that p’. We may see that it is not immediately clear clear what is meant here: if we say that someone does not know that p, we usually implicitly state that *we do know it* and that p is true. However, should we include this in our formal translation of this sentence? That is, should we formalise ‘Alice doesn’t know that p’ as ‘$\neg K_a p$’, or as ‘$p \wedge \neg K_a p$’. Ambiguities in the (implicatures of) natural language may sometimes make it difficult to decide which translation to epistemic logic is most appropriate. On the other hand, this is exactly where epistemic logic may play its role in making philosophical arguments more precise. Having attracted attention to this issue of interpreting natural language statements, we may continue to discuss the epistemic logical axioms.

Epistemic axiom (K): deductive cogency

Kripke’s Axiom, or distribution, in the epistemic interpretation also goes by the name ‘axiom of deductive cogency’. The epistemic axiom,

(K) $K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$,

reads as: ‘if agent i knows that φ implies ψ, it follows that if agent i knows φ, then she also knows ψ’. That is, our agent can be seen as being capable of doing modus ponens. Moreover, the agent instantly knows any inference of this kind. This can be seen as a form of logical omniscience: the agent is ascribed unlimited logical reasoning powers of this kind. As before, we can see that this axiom is satisfied in any possible worlds model, following from definition 2.1 of ‘K’. This incentivised philosophers to worry about the applicability of ‘possible world’-based models of epistemic logic to epistemology. One might well think of an example in which some person is unable to oversee the logical consequences of his knowledge.

Importantly, it should be noted that, as we saw before, axiom (K) is satisfied in minimal system **K** for formal reasons, and consequently valid in any standard epistemic logical model. It is a problem at the very core of epistemic logic, and will be discussed later in this chapter as a case study. There are several forms of logical omniscience. The form of logical omniscience currently under discussion is referred to by Paul Gochet and Pascal Gribomont (2006) as ‘closure under material implication’.

Epistemic inference rule (N): necessity

Like (K), the inference rule of necessitation (N) holds on every system:

[18] Imagine this to be of extreme relevance when Alice and Bob just had broken up and did not want to see each other anywhere.

(N) from φ infer $K_i\varphi$,

stating that every proposition that is valid, is known (by any agent i). As before, we can derive the validity of this principle from definition 2.1 of K_i. Postulating that the agent knows every valid proposition (or: theorem of the system) is again a case of logical omniscience. This form of logical omniscience has been called 'closure under theoremhood'. In section 2.5 we will elaborate on the several forms of logical omniscience and the philosophical discussions involved.

Nicholas Rescher holds objections against the rule of inference, because it "certainly does not hold for the standard mode of knowledge represented by K." However, he accepts the principle because it is useful, and suggests that, by accepting these kind of principles, we are not studying "knowledge proper but rather *available information*[19]." (Rescher 2005, 15)

Besides some form of logical omniscience, necessitation entails a mild kind of introspection:
If φ is some theorem of some arbitrary system, necessitation entails $K_i\varphi$. But then by definition 2.1 of the epistemic operator, $K_i\varphi$ holds everywhere and thus is a theorem of our system, therefore necessitation entails that $K_iK_i\varphi$. This is not full introspection: agent i only knows (that she knows that she knows...) that she knows all validities. However this too can be seen to diverge from any everyday understanding of knowledge: how can any human being ever know that she knows all validities, and even infinitely iteratedly know that she knows those validities?

Epistemic axiom (T): truth

Recall that system **T** is constructed by adding to **K** the axiom (T). The epistemic axiom,

(T) $K_i\varphi \to \varphi$,

reads as: 'if agent i knows φ, it is true': whatever is known is the case. That is, agents cannot have false knowledge, they cannot be wrong in knowing something. System **T** is considered to be the weakest system of epistemic interest.[20] This means that commonly in epistemic logic one investigates some strong type of knowledge, infallible knowledge: the agent cannot be wrong in knowing some proposition φ.

Rescher comes with an argument (or rather: explanation) for our acceptance of this axiom, argueing that we rationally consider our knowledge to be true:

> "Rational people are committed to seeing their knowledge as real knowledge and therefore as subject to those principles which hold for genuine knowledge in general. Now in taking our own putative knowledge to be true – that is, viewing it as actual knowledge – we accept the principle." (Rescher 2006, 480)

Although some scholars view upon (T) as an uncontroversially accepted axiom[21], some objections raised were against the truth axiom. Robert Ackermann (1972, 79) defines knowledge pragmatically as follows: "To say 'I know that p' is to assert that one can meet all relevant non-metaphysical objections to p," more specifically, that one can meet all relevant *current* non-metaphysical objections. This pragmatic notion of knowledge does not entail axiom (T), nor does it entail that (T) should not hold. Lenzen argues that the truth axiom is compatible with Ackermann's pragmatic analysis and can be consistently added to it, but it seems that this does not capture the spirit of Ackermann's philosophical analysis all too well. It seems more feasible that Ackermann is arguing

[19] Cf. 'implicit knowledge' as discussed in section 3.4.*ix*.
[20] As is contended, amongst others, by Hendricks (2005).
[21] See for example, Stalnaker 2006, 172.

for a less strong conception of knowledge: Ackermann admits that if '$\neg p$' was to be found true, one would have to give up their (pragmatic) knowledge that p. In other words, one may claim to know p until it is found to be false. We could argue that, although a person in such a case was right to *claim to know p*, he did not know it (because it was false). Lenzen (1978, 19) analyses Ackermann's words in this vein and concludes that they constitute no reason to give up (T).

Irani (1972, 492-3, cited by Lenzen 1978, 20) argues for a dismissal of the truth axiom, because it is too severe:

> "... regardless of the strength of evidence in favour of the proposition (empirical proposition), it is possible, i.e. not contradictory, to suppose that the proposition is false. From this it follows that increasing the strength of evidence ... never satisfies [the truth requirement]. Thus [the truth axiom] in the case of empirical propositions is never satisfied. Hence empirical propositions can never be known."

However, the fact that one never has ultimate evidence to prove a certain proposition p to be true, does not mean that agent i cannot truthfully believe p and have adequate evidence for it, i.e., it does not imply that agent i cannot know[22] that p . It seems that (T) is an enemy to *knowledge claims* rather than to knowledge. From an outside (God-like) perspective we might see that someone actually *knows* a certain proposition without being able to defend her knowledge claim in such a way that it satisfies the sceptic.

Moreover, Irani's discussion of the axiom seems to suggest that an acceptance of (T) implies an external, third-person point of view[23]. When an agent is not able to defend her knowledge claim, or when her knowledge claim can be defeated, the agent's subjective information may be less than the information available from a third-person perspective. Whenever the (T)-axiom is accepted and the agent is less introspective than described by the **S5**-axioms, the agent's perspective is different from the modeller's perspective. In this case, knowledge can only be ascribed from an objective, external point of view. The agent's ability to reflect upon what is true and what she knows, is further discussed in connection to axiom (4) and axiom (5).

Epistemic axiom (4): self-awareness

From system **T** we can construct system **S4** by adding axiom (4):

(4) $K_i\varphi \rightarrow K_iK_i\psi$,

which reads as: 'if agent i knows that φ, then she knows that she knows this'. This has been called positive introspection, but is also referred to as the 'axiom of self-awareness': the agent is aware of which propositions she knows.

The philosophical intuition behind this is that if you are in a strong enough epistemic position to say that you know something, you are also in a strong enough epistemic position to say that you know it, strong enough to *know* that you know it.[24] However, as is often objected: we may know things but not reflect on them. If we know something, this does not necessarily mean that we are aware of knowing it. For example, I might know that the front door of my house is green, but never really have given the matter a second thought. If you ask me what colour my door is, I will answer correctly, but before this enquiry, in pre-reflective circumstances, can I really be said

[22]In the rightly justified true belief sense of the knowing

[23]See for a more elaborate discussion on the different perspectives one can take, section 3.1.

[24]"If b is in a position to rule out all scenarios in $W - W_1$ b is *ipso facto* in a position to rule out the claim that he or she is not in such a position." (Hintikka 1986, 64)

to know that I know it?

It becomes clear that the acceptability of the current axiom is again highly sensitive to various possible characterisations of knowledge. If we have an everyday life conception of knowledge correlated to awareness, then this axiom may be contested, especially because its iteration leads to an infinite regress of self-reflection: agent i knows that she knows that she knows ... φ. If, however, we have a stronger characterisation of knowledge, the axiom of self-awareness might be defendable. Hintikka (1970, 145) points out that philosophers often maintain this stronger kind of interpretation of knowledge, and uses it as an argument for accepting the axiom of self-awareness:

> "The concept of knowledge used by many philosophers seems to be a strong one on which one knows P only if ones evidence for P is conclusive in this sense. It is plausible that the KK principle[25] holds for this strong concept of knowledge."

Earlier (1962, 84) he argued that instrospection is implied *epistemically*, in contrast to '*virtually*', or: logically. However, twenty pages later he proves the KK principle to be valid in his framework.

However, one can also easily come up with arguments for why this axiom should not hold. One example of this is that, by rejecting this principle, one can avoid the paradoxicality in the 'surprise examination paradox'. In this paradoxical story a teacher tells his students that there will be an exam next week, but that it will come as a surprise. By a simple backward induction argument the students can infer a contradiction, and can therefore indeed be surprised by the teacher. Quine's solution to this is the rejection of positive introspection which causes the backward induction argument to fail. If the exam has not taken place yet on Thursday, then, although the students might know on Thursday evening that the exam will be on Friday, they might not be aware of this. This is in keeping with the teacher's statement, so when the students start reasoning after the teacher statement, their backward induction argument does not lead to a contradiction: there is a way in which the teacher's words may come true. Therefore, by rejecting axiom (4) we may avoid the surprise examination paradox.[26]

All in all, philosophically the last word on positive introspection has not been spoken yet, but all Kripkean systems currently considered for modelling knowledge accept this axiom.

Epistemic axiom (5): wisdom

To get an even stronger notion of knowledge, we add axiom (5) to **S4**, giving us system **S5**. Axiom (5),

(5) $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$,

reads as: "if agent i does not know φ, then she knows that she does not know it." This has been called negative introspection, and referred to as the 'axiom of wisdom'. Indeed, one is truly wise if one knows of (is aware of) those things that one doesn't know. In system **S5**, combining axiom (4) and (5), the agent has full access to the epistemic alternatives (the accessibility relation is an equivalence relation: reflective, symmetric and transitive). This means that the concept of knowledge considered in this system is very strong.

The axiom of wisdom can be compared to the closed-world assumption that is used in computer science. A database of some airline company, for instance, satisfies the closed-world assumption if from there not being an entry with a direct flight from Amsterdam to Ulan Bator, Mongolia, it

[25]'KK principle' is yet another name for axiom (4).

[26]For more on this, see the elaborate discussion of the surprise examination paradox in Kasbergen 2012.

can be inferred that there is no such flight available. This condition is not very realistic if one talks about human beings. However, it is commonly assumed for programs, robots and automata. The strong system **S5** is taken to be the right system for modelling knowledge by computer scientists Fagin, Halpern, Moses and Vardi (1995) in their influential book *'Reasoning about Knowledge'*.

In ordinary life, however, we do not have such privileged access to our epistemic state. This is for Lenzen (1978, 79) a reason to dismiss the axiom of wisdom as "unacceptable." In same vein, Stalnaker argues:

> "it is hard to see how any theory that abstracts away from the possibility of error could be relevant to epistemology ... I think that the possibility of error, and the differences between knowledge and belief are relevant to the intended domains of application of ... models [of computer scientists too]." (Stalnaker 2006, 177)

Furthermore, Hintikka (1962, 54) rejects the axiom of wisdom, because "[t]he consequences of this principle ... are obviously wrong." Hintikka proves that in the system he developed, axiom (5) implies '$p \to K_i P_i p$', i.e. in our notation '$p \to K_i \neg K_i \neg p$'. To Hintikka it is "perfectly obvious," that the principle 'if p is true, then the agent knows that it is possible that p' is "not acceptable."

In contrast to these rejections of axiom (5), others see it as a useful assumption related to certain applications:

> "Despite these problems, the **S5** properties are a useful idealisation of knowledge for many applications in distributed computing and economics, and have been shown to give insight into a number of problems. The **S5** properties are reasonable for many ... examples." (Van Ditmarsh et al. 2015, 36)

An explicit example that is given in the *'Handbook of Epistemic Logic'* is that of an interpreted system, a system that describes the states of some computer process (for instance, two processors computing three variables x, y and z). Its relation is an equivalence relation, therefore in this system all the knowledge axioms so far discussed hold. Thus, an interpreted system satisfies all the knowledge axioms, and therefore **S5** is a good way of modelling it.

In search for a right modal logic to capture knowledge, several suggestions have been done for systems that lie between **S4** and **S5** in strength. Table 2.2 reviews the systems that are commonly discussed and their axioms. The upward arrows in table 2.2 signify increasing strength.

Table 2.2 Modal systems of epistemic interest (Based on Hendricks 2005)

System	**Axioms**	**Strength**
S5	**K** + (T) + (4)	
S4.4	**K** + (T) + (4) + (4.)	↑
S4.3	**K** + (T) + (4) + (3.)	↑
S4.2	**K** + (T) + (4) + (2.)	↑
S4	**K** + (T) + (4) + (5)	↑

Epistemic axiom (.2)

System **S4.2** results from system **S4** by adding axiom (2.),

(2.) $\neg K_i \neg K_i \varphi \to K_i \neg K_i \neg \varphi$,

which can be intuitively read as follows: 'if you consider it epistemically possible that you know φ, then you know that you consider φ epistemically possible'. Lenzen claimed that a logic of knowledge should be at least as strong as system **S4.2**, and at most as strong as system **S4.4**. To arrive at this conclusion he considered two possible ways of defining knowledge: knowledge as true strong belief, and knowledge as a more demanding concept described by several philosophically defended axioms. (Lenzen 2004)

Epistemic axiom (.3)

System **S4.3** results from system **S4** by adding axiom (3.),

(3.) $K_i(K_i\varphi \rightarrow K_i\psi) \vee K_i(K_i\psi \rightarrow K_i\varphi)$,

and is advocated by Wiebe Van der Hoek (1993) to be the right system for modelling knowledge. Also Robert Stalnaker (2006, 187-91) argued for system **S4.3**, because it captures a definition of knowledge as true justified belief plus the requirement that there be no "defeater." The 'safe belief'-operator of Alexandru Baltag and Sonja Smets (2016), meant to express 'defeasible knowledge', also satisfies the the axioms of **S4.3**.

Epistemic axiom (.4)

System **S4.4** results from system **S4** by adding axiom (4.),

(4.) $\varphi \rightarrow (\neg K_i\neg K_i\varphi \rightarrow K_i\varphi)$,

which can more intuitively be read as: 'if φ is the case, and you consider it epistemically possible that you know φ, then you know φ. This principle is valid if you take knowledge to be true, strong belief, as is advocated by Franz von Kutschera. (Von Kutschera 1976) Strong belief in this sense means belief of such a high degree, that it cannot be lost. Defining knowledge in these terms enables an agent to maintain wrong beliefs about what she knows, that is, the axiom of wisdom fails.

2.4 Examples of idealisations in epistemic logic

As we have seen from the previous section, there are many ways in which epistemic logic deviates from the reality of human knowledge. In the epistemic logical community it is common to refer to these deviations from reality as 'idealisations'. As we will see later in chapter 4 from a literature study of the philosophy of science, the terms 'idealisation' and 'abstraction', which are supposed to capture different ways of deviating from reality[27], have many different possible definitions. To capture any such deviation therefore, we are safer to use the collective noun 'representational imperfection'. Jones (2005) suggested the use of this word. It is meant to be a neutral word to capture any deviation from the reality of the target system. However, this is not in line with common usage in epistemic logic, and therefore may be confusing. For that reason, we choose to speak of 'idealisation' in this and the following chapter whenever we want to refer to any kind of representational imperfection. In chapter 4 the definitions of all these concepts will be examined, leading to a clear redefinition of the concepts involved which can thereafter be used.

There are many different kinds of idealisations in epistemic logic. This section will give some examples of them, in order to excite the imagination. A full list is not possible, as there are (infinitely) many idealisations because of the complexity of nature. The examples of idealisations will be discussed along four categories.

[27] More generally, deviating from the target system.

Idealisations related to the axioms

Firstly, we note that some idealisations are connected to some particular system(s) of modal logic. These idealisations have been elaborately discussed in section 2.3.

Qualitative idealisations

The representational imperfections that often first come to mind when you ask a non-logician, are of a material or qualitative nature, and present in any model (based on modal logic) because of its mathematical nature. Examples of this would be personal details about the agents: The agent loses all her personal details (e.g., name, nationality, ethnicity, language she speaks, sex and gender, et cetera), all her character traits (e.g., her being oblivious, easily upset, contemplative, credent, et cetera) and most of her personal history (e.g., good and bad experiences she had, learned behaviour, approaches and strategies, and most of the learned information, also –perhaps unconscious– ideas from the religion, profession and cultural group she belongs to). Furthermore, she loses her embodiment and her present state of mind (she can be happy, sad, in love, drugged, moody, et cetera).

Some of such details about the agent will be considered to be relevant for her epistemic state (e.g., her being oblivious or not being able to apply some logical rule), while others will be considered to be irrelevant (e.g., hair colour[28]). Naturally, what is deemed relevant or irrelevant depends on the context of the modelling situation. For example, if you want to model communication between people of different cultures realistically, we should include learned language behaviour in our models. When an English boss tells a Dutch employee that something is 'no problem at all', the employee might understand from this that there is no problem at all. However, a British bystander might understand from the situation that there is indeed a very grave problem, and that the boss was just speaking politely. Cultural subtleties like these do, in very many other applications, not play a role. What is deemed relevant depends on the task at hand.

Useful fictions

Some idealised concepts are consciously introduced into an epistemic logical model. An example of this is the concept of 'common knowledge', which is amply used in epistemic game theory, for instance. A group of agents G have common knowledge of φ iff[29] all agents in G know φ, they all know that they know φ, they all know that they know that they know φ, and so on ad infinitum. The concept of common knowledge makes use of infinite knowledge iterations, which is a deviation from human reality as human agents do not entertain such infinite knowledge iterations in their mind.[30] Nevertheless, for many applications this can be a very useful fiction.

Other byproducts of our formalisation

There are some idealisations that were not intended when constructing a model, but turned out to be byproducts of our formalisation. Examples of this would be the several forms of logical omniscience. In the next section we will discuss logical omniscience as an elaborate case study. It is useful to look at this issue more technically, and compare different solutions to the problem of logical omniscience. By doing so we get a more thorough understanding of the origin of this idealisation and why it is so difficult to get a good logical model with realistic properties. Also it is important to perform some epistemic logical work ourselves, instead of only analysing works

[28]Unless one believes dumb blonde jokes.

[29]If and only if.

[30]However, a group of agents can have common knowledge of some formula, for example, after a public announcement of that formula. See for instance (Baltag, Moss and Solecki 2016).

and researchers in epistemic logic; when we get our hands 'dirty' this way, we get a better grasp on the field of epistemic logic.

2.5 Case study: logical omniscience

Logical omniscience is one of the most persistent and most discussed idealisations of epistemic logic. In its most common form, logical omniscience is described as follows: the agent immediately knows all the logical consequences of her knowledge (in particular, she knows all the theorems of the system). This is a clear deviation from actual (human) agents, and it is often regarded to be undesirable. We therefore speak of the 'problem of logical omniscience'. However, there is not just one problem of logical omniscience. Rather, there are a family of logical omniscience problems. Gochet and Gribomont (2006, 99) listed seven forms of logical omniscience:[31]

Definition 2.2. *(Forms of logical omniscience)*

1. *If* $\vdash \varphi$ *then* $\vdash K\varphi$ *(closure under theoremhood);*
2. *If* $\vdash \varphi \to \psi$ *then* $\vdash K\varphi \to K\psi$ *(closure under logical implication);*
3. *If* $\vdash \varphi \equiv \psi$ *then* $\vdash K\varphi \equiv K\psi$ *(closure under logical equivalence);*
4. $\vdash K(\varphi \to \psi) \to (K\varphi \to K\psi)$ *(closure under material implication);*
5. $\vdash K(\varphi \equiv \psi) \to (K\varphi \equiv K\psi)$ *(closure under material equivalence);*
6. $\vdash (K\varphi \wedge K\psi) \to K(\varphi \wedge \psi)$ *(closure under conjunction);*
7. $\vdash K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$ *(closure under simplification).*

Not all forms of logical omniscience are considered to be equally objectionable. For example, one could argue that our agent is logically strong enough to perform closure under simplification. The last four forms of logical omniscience have at times been found defendable. However, the first three forms of logical omniscience, are said to be "really unwanted." (Gochet and Gribomont 2006, 158) The idea behind this could be that the agent does not have privileged access to the logical truth of the system she finds herself in.

Now let us have a closer look at the seven forms of logical omniscience. For brevity , the defined forms of logical omniscience will sometimes be referred to as LO(i) for i the number in the list of Gochet and Gribomont. It can be easily seen that LO(1) is equal to the inference rule (N) of standard epistemic logic, which was discussed two sections ago. Likewise, LO(4) is equivalent to axiom (K). Without any further reasoning, we can thus conclude that at least these forms of logical omniscience will occur in all normal modal logics[32]. Also the other forms of logical omniscience can be checked to hold in all normal modal logics. We provide two instructive examples in Appendix B.

2.5.1 Solutions to logical omniscience

Many solutions have been raised to the logical omniscience problem. The discussed approaches are instructive of how one can adapt, specify or generalise standard epistemic logics in order to escape logical omniscience. These different approaches attack different reasons for why actual agents fail to be logically omniscient. Therefore, we will first have a look at possible sources of

[31] As we restrict to a single agent, we omit the index i.

[32] Normal modal logics are modal logics that satisfy axiom (K) and inference rule (N). See Appendix A for an introduction to normal modal logics.

failure of logical omniscience for real agents.

We distinguish three different kinds of sources:

1) *Limited logical ability.* Real agents are not infallible: we have an imperfect understanding of the logical rules, or may make mistakes in their reasoning.

2) *Limited resources.* Often when talking about the problem of logical omniscience, researchers refer to resource-bounded agents: real agents have insufficient time, memory or computational power needed to follow through all the consequences of their knowledge.

3) *Psychological sources.* The logical omniscience problem has inspired a solution in terms of awareness. Other solutions have been couched in terms of prejudices, attention entrenchment and little importance as sources of lack of logical omniscience. (Huang and Kwast 1990)

Syntactic logic One way of dealing with logical omniscience is the syntactic approach[33]. The intuition behind this approach is that we can, for every possible world w, list all the formulas that the agent knows at w. In this approach there is no reference any more to the original way of defining knowledge as is common in epistemic logic: knowledge as holding in all accessible possible worlds. By simply listing all the formulas that an agent knows at a certain world, there is no need any more for the accessibility relation. The internal dynamics[34] of the syntactic structure, therefore, are also less interesting than usually in possible world semantics. Although clearly all forms of logical omniscience can be avoided in this way, we may object that the syntactic approach merely gives us knowledge representation, and does not show us anything interesting about the concept of knowledge. The agent is in this approach very static and logically blind: she cannot perform any logical inference whatsoever. We could equip the agents with some more logical abilities by forcing constraints on the syntactic knowledge function $\mathcal{C}(w)$.[35] However, instead of going into detail here, we will look at a weak modal logic that represents such limited (but not completely lacking) logical abilities.

Weak modal logic Equipping the agent with a bit more, but still limited, logical abilities, Andrzej Wiśniewski (1998) constructs two weak modal logics of occurrent belief. In these logics, the limited logical abilities of an agent are mimicked by an incomplete set of logical rules that the agent has at her disposal. We will discuss the first of Wiśniewski's systems: **S.0**. Its axioms are all the tautologies of the classical propositional language, together with closure under material implication LO(4), closure under conjunction LO(6), closure under simplification LO(7) and the consistency axiom (not pertaining inconsistent knowledge).

The agent of Wiśniewski has some logical abilities: because of the consistency axiom she will not know[36] contradictions. However, this does not mean that she herself knows the law of non-contradiction: she might well fail to see this tautology of the classical propositional language. It can be seen that system **S.0** avoids the first three forms of logical omniscience. These forms are said to be the most unwanted forms of logical omniscience; a positive point for the approach of weak modal logics of Wiśniewski.

[33]Discussed as in (Fagin et al. 1995).
[34]See section 4.1.4 for a more elaborate discussion of internal dynamics.
[35]It is minimally defined to satisfy classical propositional logic for the atoms and connectives.
[36]Wiśniewski calls this belief.

Awareness logic Fagin and Halpern (1987) proposed to model failure of logical omniscience by means of an additional awareness function. The syntactic awareness function maps each world to a set of propositions that the agent is aware of at that world. Knowledge is defined as truth among all accessible worlds, as is standard in epistemic logic, combined with the requirement that the proposition is also in the awareness set of the agent at that world.

Now we can have a look at which forms of logical omniscience get avoided by adding the awareness function. Naturally, this depends on the features of the awareness function. In principle, by means of the awareness function all forms of logical omniscience can be avoided. However, we can strengthen the agent's logical power by imposing conditions on the awareness sets belonging to the possible worlds. Suppose, for instance, that the awareness set of the actual world is upwards closed. This means that if all subformulas of some compound formula φ are elements of the awareness set, then also φ is an element of the awareness set. Then it can be seen that closure under conjunction, LO(6), holds again: if φ_1 and φ_2 are in the awareness set, then the fact that it is upwards closed entails that $\varphi = \varphi_1 \wedge \varphi_2$ is in the awareness set. So if we know φ_1 and we know φ_2, then it quickly follows that we know the compound $\varphi = \varphi_1 \wedge \varphi_2$.

Similarly, it can be checked that if the awareness set of the actual world has the property that all the subformulas of each formula in the awareness set are also elements of the awareness set, that is, if the awareness set is downwards closed, then logical omniscience forms LO(4), LO(5) and LO(7) persist. These examples show the tight connection between the properties of the awareness function and the forms of logical omniscience that persists. Naturally, when we take our awareness function to map every world to the set of all propositions, we again are back with standard epistemic logic and inherently all forms of logical omniscience.

Now, when the awareness set of a world is smaller than the set of all propositions, we usually avoid the first three forms of logical omniscience: we cannot know a theorem unless we are aware of it, closure under logical implication fails if we are unaware of the consequent, and we cannot know two formulas to be logically equivalent if we are unaware of (one of) them. This must surely be a point in favour of the awareness approach, because as Gochet and Gribomont discussed, the first three forms of logical omniscience are the most objectionable. However, the awareness approach is rather one-sided: it only enables us to model failure of logical omniscience due to lack of awareness. As discussed in the previous paragraph, there are more sources of failure to be logically omniscient.

Justification logic As we may recall from the epistemological discussion of different definitions of knowledge, knowledge is commonly required to be justified (in the right way). Whereas standard epistemic logic does not give us the tools to model this, justification logic enables us to speak about the evidence that we have for knowing a certain proposition. Adding justifications to the language even enables us to escape Gettier's example of the red barns[37], as suggested by Sergei Artemov and Melvin Fitting (2016).

Most justification logics are assumed to satisfy the logical awareness principle, which means that the agent accepts all logical axioms as justified. Of course, this leads to the first form of logical omniscience, closure under theoremhood LO(1). However, we can avoid this by letting our constant specification[38] reflect the agents logical acumen.

Placing certain conditions on the constant specification of our justification logic determines what logical powers our agent has. If we define our constant specification to be 'axiomatically appropri-

[37]See section 2.1.1 for a discussion of the definition of justified true belief and Gettier's counterexamples (1963).
[38]See appendix B for an explanation of the terminology used in this section.

ate', it can be proven that this validates internalisation. (Artemov and Fitting 2006) This means that, when the constant specification of our justification logic is axiomatically appropriate, our agent is logically omniscient in the sense of knowing all the theorems, LO(1). When the constant specification is empty, however, we have a completely sceptical agent who has no evidence for any axiom. Slightly less extreme, we may want to define our constant specification to be finite. This will mean that the agent has justifications for some occurrences of (some of) the axioms, but fails to be logically omniscient in the sense of LO(1). Similarly, imposing constraints on our constant specification enables us to avoid LO(2) and LO(3) as well.

Now let us have a look at closure under material implication, LO(4). We interpret this as follows: at any model and any world, if we have conclusive[39] evidence for $\varphi \to \psi$, and if we have conclusive evidence for φ, then we have conclusive evidence for ψ. So suppose, for arbitrary possible world justification model M and world w that we have justification t such that $M, w \models (t : \varphi \to \psi)$, and we have justification s such that $M, w \models (s : \varphi)$. Now the question is: do we then have some justification r such that $M, w \models (r : \psi)$? The answer is yes. By our application axiom we find that, applying t to s gives us conclusive evidence for ψ. So for $r = t \cdot s$ we find that $M, w \models (r : \psi)$. It follows that, because the application axiom holds in all justification logics, justification logics are closed under material implication.

The other forms of logical omniscience do not hold generally, we can find counterexamples in which they fail. Let us for instance look at closure under conjunction. Consider a possible worlds justification model M consisting of just one world w with a reflexive accessibility relation. Suppose that $M, w \models (t : \varphi)$ and $M, w \models (s : \psi)$. This means that w satisfies both φ and ψ, and that $w \in \mathcal{E}(t, \varphi)$ and $w \in \mathcal{E}(s, \psi)$. Because we had classic propositional logic underlying justification logic, it follows that, in all worlds accessible from w, $\varphi \wedge \psi$ holds. However, we might well think of an evidence function $\mathcal{E}$ such that $\mathcal{E}(r, \varphi \wedge \psi) = \emptyset$ for any term r. Such an admissible evidence function provides us with a counterexample, so we conclude that justification logics are not generally closed under conjunction.

Adapting our constant specification and putting constraints on our admissible evidence function give us a possibility to speak of a less than logically omniscient agent. Moreover, Artemov and Kuznets (2009) prove that justification logical agents are in general less logically omniscient than epistemic logical agents. They do so by redefining logical omniscience in terms of computational complexity and defining a test of this accordingly. Their definition states that an agent that is able, for each formula φ that she knows, to give a proof of φ in polynomial time in the size of it[40], is called *not logically omniscient*. This is quite a drastic redefinition, but the discussion given by Artemov and Kuznets (2009) shows a difference in complexity of proving knowledge statements between justification logic and (standard) epistemic logic. However, a more elaborate discussion of this will be outside the scope of this thesis.

Impossible worlds semantics In reaction to the critique of logical omniscience, Hintikka (1979) suggested that we may relax the notion of possible worlds. Naturally, when we allow our agents in our models to only consider logically possible worlds, they will be logically omniscient. Therefore, if we want to avoid logical omniscience in our models, we only need to include the option that the agent considers some logically impossible worlds possible. A world may be *epistemically possible* from the agent's point of view while being *logically impossible* (from the modeller's point of view). What exactly impossible worlds are is debatable[41], however, we will take them to be

[39]That is, evidence that leads to knowledge. In this terminology we follow Artemov and Fitting 2016.
[40]That is, in the size of the knowledge assertion. This will be '$t : \varphi$' for justification logics and '$K\varphi$' for epistemic logics.
[41]See Berto 2013.

possible worlds in which the laws of the logic chosen by the modeller fail. Of course, whether some world is logically impossible or not depends on the choice of logic that underlies your model: for example, a classically impossible world may be an intuitionistically possible world. For now, we choose the logic of modeller to be classical logic, the logic that underlies standard epistemic logic. However, we should remain aware of the fact that we could make a different choice here.

We refer to the current type of non-normal modal logics as 'impossible worlds semantics'. To illustrate how impossible worlds semantics work, we look at the example in Figure 2.1.

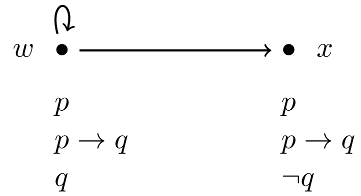

Figure 2.1: Impossible worlds model

In this example w is a *possible* possible world, whereas x is an *impossible* possible world. The (classical) illogicality of world x causes a failure of logical omniscience in world w. Because we have that $R(w, w)$ and $R(w, x)$, it follows that world w satisfies Kp and $K(p \to q)$, but $\neg Kq$. That is, closure under material implication, LO(4), fails in world w. Similarly, we can construct examples in which the other forms of logical omniscience fail. Let us have a look at, for instance, closure under logical equivalence, LO(3). This form of logical omniscience is hard to escape in epistemic logic, because standardly propositions are defined in terms of sets of (logically) possible worlds. Because of this focus on propositions rather than on sentences, the agents cannot distinguish between sense and reference[42]: after all, if two sentences refer to the same proposition, the sets of possible worlds in which they are satisfied are exactly the same. However, impossible worlds semantics is not necessarily closed under logical equivalence. Let φ and ψ be two formulas that are logically equivalent: $\varphi \equiv \psi$. Note that because of the introduction of logically impossible worlds this does no longer mean that the truth sets of φ and ψ are the same: we may have an impossible world x such that $M, x \models \varphi \wedge \neg\psi$. If this impossible world is accessible for the agent, she may know φ while failing to know ψ, even though φ and ψ are logically equivalent.

In Example 1 it was easy to see that x was a logically impossible world. However, not all logical flaws are so easy to spot. Many real agents will be able to see the failure of a logical law in $p \wedge (p \to q) \wedge \neg q$, but this is not so easy for a formula like $\neg\neg p \wedge \big(\neg(q \vee (\neg q \wedge (p \to p))) \vee \neg p\big)$. We might want to model an agent that can spot failures of logical laws that are easily seen, but that has trouble with seeing through logical failures in more complex formulas. This can be done, as Hintikka (1979) suggested, with the help of Rantala's 'urn models' (Rantala, 1979). In these urn models, a certain parameter d represents the agents logical acumen. In this approach, logically possible worlds are construed as invariant urn models while impossible possible worlds are construed as invidiously changing urn models. An impossible world can by the agent be considered to be epistemically possible when it varies only imperceptibly: the model looks like an invariant model to the agent because the logical acumen of the agent is not high enough to understand the subtle illogicality. Let us call a d-invariant model an urn model that is invariant until modal depth d. Then an agent with logical acumen level d will always have it as an epistemically possible world, whereas agents with a higher level of logical acumen may not. Of course, a d+1-invariant model is always a d-invariant model: that what a stronger agent accepts as epistemically possible, will also be accepted as epistemic possible by an agent with lower logical acumen. Taking this

[42]For more information on the sense-reference distinction we refer to (Frege 1980).

approach we can model lack of logical omniscience because of having limited logical abilities and because of lacking resources.

Neighbourhood semantics Generalising modal logics by introducing neighbourhoods may also give us a way out of logical omniscience. In neighbourhood semantics[43] the agents knowledge at a certain world is identified with a neighbourhood by means of the neighbourhood function, $N(w)$. For our present purpose we interprete the neighbourhood function epistemically as giving us the known sentences in a world:

$M, w \models K\varphi$ iff $||\varphi||_M \in N(w)$

To illustrate how this works, we have a look at the example in Figure 2.2:

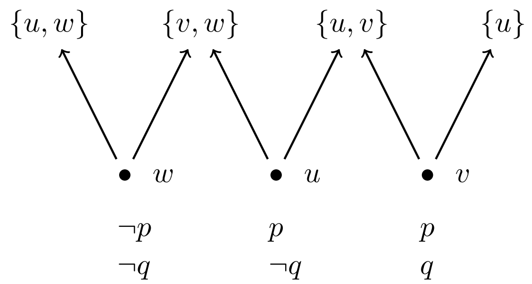

Figure 2.2: Neighbourhood model

In this example the truth set $||p||_M$ of p is given by $\{u, v\}$. By the definition of knowledge for neighbourhood semantics just given, we know that for all models M and worlds $w \in W$, $M, w \models Kp$ iff $||p||_M \in N(w)$. From the illustration it is easy to see that this holds for worlds u and v, and not for world w: $\{u, v\} \in N(u)$ and $\{u, v\} \in N(v)$, whereas $\{u, v\} \notin N(w)$. So we may conclude that $M, u \models Kp$ and $M, v \models Kp$, but $M, w \not\models Kp$. Furthermore, we can see that $||p \to q||_M = \{v, w\}$ and $||q||_M = \{v\}$. So by the definition of knowledge, it holds that $M, u \models Kp$, $M, u \models K(p \to q)$ while $M, u \not\models Kq$. That is, closure under material implication, LO(4), fails at world u.

Likewise, we can come up with examples in which other forms of logical omniscience fail, except closure under logical equivalence, LO(3). This is the case because, as earlier discussed, propositions are defined in terms of their truth sets, and the neighbourhood function ranges over sets of worlds. Let us suppose that φ and ψ are logically equivalent: $\varphi \equiv \psi$. By definition this means that $||\varphi||_M = ||\psi||_M$. Then clearly for all models $M = (W, N, ||\cdot||)$, and worlds $w \in W$, $||\varphi||_M \in N(w)$ if and only if $||\psi||_M \in N(w)$. So φ is known if and only if ψ is known, and we conclude that closure under logical equivalence always persists in a neighbourhood model. Consequently, whenever the agent knows some really simple tautology, by closure under logical equivalence this implies that the agent knows all axioms of the system. So in virtually all neighbourhood logical models that we want to use for modelling (realistic) epistemic scenarios, our agent is logically omniscient in the sense of LO(1) and LO(3). All other forms of logical omniscience can be avoided.

2.5.2 Discussion solutions to the logical omniscience problem

There are many different reasons why an agent may fail to be logically omniscient. In the same way, there are many different solutions to the problem of logical omniscience. Many of the approaches have been motivated to solve the problem of logical omniscience with a particular source

[43]Discussed as in (Pacuit 2007).

of the failure of logical omniscience in mind. For instance, awareness logic has been constructed in order to capture the lack of logical omniscience due to lack of awareness of some proposition(s), and impossible worlds semantics is meant to capture the fact that real agents sometimes think logically impossible[44] situations to be possible.

The syntactic approach and weak modal logics both give us a way to model limited logical abilities of the agent because of knowing fewer logical rules. In the first approach we may avoid all forms of logical omniscience, on the price of being completely logically blind. In the second approach we are able to model an agent with more logical abilities, while still avoiding the 'worst' forms of logical omniscience, LO(1)-LO(3). Another way of modelling the agent's limited logical abilities is allowing logically impossible worlds to be deemed epistemically possible. In impossible worlds semantics we can avoid all forms of logical omniscience: the agent can make mistakes. A bit more realistically even, we can make the logical failures depend on a parameter that describes the agent's logical acumen.

Impossible worlds semantics attract attention to the fact that our underlying logic does not capture the way that humans reason well. Classical propositional logic underlies standard epistemic logic, and many of its adaptations as discussed in this section. Look for example at awareness semantics, justification logics and neighbourhood semantics: for all of them, their possible worlds are commonly assumed to satisfy classical propositional logic. The way they avoid certain forms of logical omniscience is by adding extra conditions to the definition of knowledge. This issue will be discussed further in chapter 5.

Besides modelling the agent's limited logical abilities, impossible worlds semantics also enable us to model the lack of logical omniscience due to limited resources. The only other discussed approach that is intuitively capable of doing this is the syntactic approach, although as discussed it can only model no resources at all. For other approaches to modelling limited resources we refer to more quantitative solutions to the logical omniscience problem, like those discussed in (Fagin et al. 1995).

A quite different reason for the lack of logical omniscience can have a psychological origin, like the lack of awareness. This is targeted by awareness semantics, which enable us to avoid the first three forms of logical omniscience. In this way we can again give the agent more or less power (for example, making her aware of all subformulas of the propositions that she is aware of), concluding in possibly keeping some of the last four forms of logical omniscience. A major downside of the awareness approach is that it can only model lack of logical omniscience stemming from a psychological source.

An approach that seems to be related to the awareness approach is justification logic. Recall that knowledge of a proposition in awareness models requires the standard Hintikka/Kripke clause[45] plus being aware of the proposition. Similarly, knowledge of a proposition in justification logics requires the standard Hintikka/Kripke clause plus having admissible evidence for the proposition. Now the main difference here is that justification logic has additional operators on its justifications and thereby additional closure conditions on its evidence function. Therefore, it has been suggested that possible world justification models can be regarded as a dynamic version of awareness models. (Artemov and Fitting 2006) Justification logics have the additional attractive feature that they give us tools to model evidence and justifications that we might have for sentences. Thereby we are able to avoid the first three forms of logical omniscience easily, which is a major

[44]It should be remembered that what counts as logically impossible depends on the chosen logic. The agent does not necessarily reason by this logic.

[45]The requirement that all accessible worlds satisfy the proposition in order for it to be known.

advantage for many applications. At the same time, it is intuitive and connects to the definition of knowledge as (rightly) justified true belief discussed in epistemology.

Neighbourhood semantics enable us to avoid all but the third form of logical omniscience, although LO(1) gets smuggled in again as a possible consequence of LO(3). This is bad news for neighbourhood semantics as a solution to logical omniscience, because these forms of logical omniscience are considered to be truly objectionable. The motivation for neighbourhood semantics was rather technical, and it is for philosophical purposes a bit less useful because it is quite unintuitive how to model a concrete situation best with the help of neighbourhood semantics. As neighbourhood semantics are a generalisation of (normal) possible worlds semantics, we can construct a neighbourhood model from a standard epistemic logical model.[46] However, if we do this we again end up with all the forms of logical omniscience.

There is one more connection here that is easily overlooked: the connection between the syntactic approach and neighbourhood semantics. In the first we listed at each world simply the formulas that the agent knows, in the second we listed at each world its neighbourhoods: sets of possible worlds representing semantically the formulas that the agent knows. Whereas in the syntactic approach all forms of logical omniscience got avoided, in neighbourhood semantics closure under logical equivalence remains because of the representation of the agents information in terms of sets of possible worlds. However, both are knowledge representation rather than telling us anything interesting about the concept of knowledge. Compare this to standard epistemic logical models, that explain knowledge in terms of the accessibility relation.

2.6 Discussion

Idealisations in epistemic logic are often considered to be problematic. It is not for nothing that the topic of our case study in the previous section is often referred to as the *problem* of logical omniscience. The various forms of logical omniscience are pervasive in both standard and non-standard epistemic logic. There seems to be a conflict between wanting to model agents that have some rationality or logical ability, and the fact that even very complex logical inferences consist of finite (simple) inference steps. As such, any (qualitative) approach will be a trade-off between the one and the other. Some approaches seem to have a better balance than others, but which choice of logic will be best, depends on the application that one has in mind. Some of the approaches may be too detailed or too complex to use, others may be not comprehensive enough, perhaps failing to capture lack of logical omniscience stemming from different sources. As Joseph Halpern's words[47] show, this can be an incentive to look for other solutions:

> "None of the methods [dealing with logical omniscience] we have now is as good as I would want. I think there's more to be done. It seems that the best method is somewhat application-dependent, I don't think there's any really great methods out there."

Whether or not some idealisation is conceived of as a problem depends on the research context. In order to assess whether the idealisations inherent to epistemic logic are acceptable or not, we need to take into account the intended aims of epistemic logic, the standpoint with respect to normativity or descriptivity, the intended conception of knowledge to be studied, et cetera. To get a better idea of what these aims and standpoints are, in the next chapter we will have a look at the opinions and considerations of epistemic logicians currently working.

[46]By taking the neighbourhood function of each agent i at each world w, to be the set of supersets of the set of all 'i-neighbours' ($R_i(w) := \{w' \mid wR_iw'\}$) of that world in the standard epistemic logical model.

[47]This citation stems from the interview conducted with Halpern for this thesis project.

Chapter 3

The Practice of Epistemic Logic

In this chapter we will discuss what researchers say about the practice of epistemic logic. As became clear in the last sections of the previous chapter, we cannot really say to what extent some idealisation is problematic, without referring to a particular intended aim, application and context. However, aims and success criteria of epistemic logic are rarely discussed in the literature. Hintikka notes, "the technicalities of epistemic logic ... routinely receive the lion's share of attention in books and papers on epistemic logic." (Hintikka 2003, 33) The aim of the practice of epistemic logic, and its philosophical justification, usually appear just as one or two sentences somewhere hidden in the middle of a large work, easily overlooked, or even remain only implicit. In order to fill this gap, we asked 10 researchers from different research fields about their opinions regarding the subject matter. This material is combined with a study of the literature, among which the book *'Epistemic Logic: 5 Questions'*, which voices the opinions of 25 epistemic logicians on similar subject material.

Before we can have a look at what the researchers have said concerning the practice of epistemic logic, we first have a look at the differences that can exist between the works of different researchers, and the interconnectivity between these issues. Here is a non-exhaustive list of issues that should be addressed (conscious or not) when modelling knowledge: the aim of your research, the extent to which your research is descriptive or normative, the object of study, the way knowledge is construed to be, how strong we take knowledge to be, knowledge of propositions versus knowledge of sentences, knowledge claims versus knowledge attribution, first versus third-person perspective, et cetera. Although most of these issues have been introduced in the previous chapter, we turn to a more elaborate discussion of some of them and their interconnectivity before we present the outcomes of the interviews we conducted.

3.1 Several topics of debate in epistemic logic

Descriptive versus normative (reappraisal) This issue poses whether epistemic logic represents real human knowledge *as it is*, or if it rather determines how we *should* or *ought to* reason about knowledge. The first standpoint we call descriptive, the second normative. Of course, a model can be partially descriptive and partially normative. As we have seen from our discussion in section 2.1.2, in epistemology, purely descriptive and purely normative approaches are hard to come by.

Hintikka (1962, 38) remarks the following with respect to the normativity of (epistemic) logic:

> "The fact that the so-called laws of logic are not 'laws of thought' in the sense of natural laws seems to be generally admitted nowadays. Yet the laws of logic are not laws of thought in the sense of commands, either, except perhaps laws of the sharpest

> possible thought. Given a number of premises, logic does not tell us what conclusions we ought to draw from them; it merely tells us what conclusions we may draw from them–if we wish and if we are clever enough."

Although Hintikka considers epistemic logic, or logic in general, to *not* be a normative practice, we cannot deny that there is a normative feel to his words: perhaps logic could provide "laws of the sharpest possible thought." Also Lenzen's definition of the task of epistemic logic has a normative feel to it: "elaborating the logical laws which one may *rationally expect* the belief- and knowledge-system of the subject *a* to obey." (Lenzen 2004) Although these authors may deny the full normativity of epistemic logic, their wording in terms of 'laws of the sharpest possible thought' and 'laws that one may rationally be expected to obey' are confusing the message.

A normative construal of epistemic logic faces some problems. For example, how do we interpret axiom (T), and can it really be a norm for us humans to have infinite rationality, perfect recall, logical omniscience, et cetera. Another question that we may ask ourselves is: how can we set up rules for an ideal agent if we are not ideal ourselves?

Epistemic logic is not considered to be a fully normative practice. Neither is epistemic logic looked upon as a purely descriptive practice, because of all the unrealistic assumptions that are made. Lenzen (1978, 14) argues along these lines: "Since the possibility of 'logically obtuse men' may be taken for granted, epistemic logic must indeed not be viewed as a descriptive (empirical) theory of what people factually believe and know." Furthermore, a complete, veracious description of knowledge as in real-life situations is impossible.[1] So both the purely normative and the purely descriptive view are untenable.

Epistemic logic may not be purely descriptive or purely normative, however it may have descriptive or normative elements. We already discussed some normative elements in the words of Hintikka and Lenzen. Van Benthem (2006, 55, 57, 72) attracts attention to the descriptive elements of epistemic logic, in emphasising that common parlance and context are essential to epistemic logic. Van Benthem is not the only one that discusses descriptive elements of epistemic logic: Lenzen (1978, 15) argues that an adequate notion of knowledge should capture its "basic meaning" (which he does not further explicate), and Hintikka expresses the hope that epistemic logic has enough in common with our natural use of the term 'knowledge'. (Hintikka 1862, 10)

Although opinions differ regarding the extent to which epistemic logic is descriptive, most researchers agree that descriptivity has a role to play in epistemic logic. Knowledge is a concept that is taken from our everyday life: if we did not have the concept of knowledge in our everyday life, there would have been no need, no incentive, to construct a logic of knowledge.

The way knowledge is construed to be There are different ways in which to construe a logic of knowledge. Firstly, we may construe knowledge in various degrees of strength (accepting more or less axioms discussed in section 2.3.2). Secondly, we might choose to focus on knowledge of sentences rather than on knowledge of propositions. As we have seen from the previous chapter, this might be feasible because it gives us a way to avoid some form of logical omniscience, and it enables us to model human knowledge more realistically: if we know that we see the morning star, we might not be aware of the fact that we also see the evening star.[2] Moreover, one can choose to direct attention towards knowledge claims rather than to knowledge attribution (by an omniscient

[1] We will elaborate on this in section 3.4.*ii*.

[2] The name 'morning star' is referring to the same planet (Venus) as the 'evening star', however for centuries people have not been aware of this fact. It is a locus classicus of the sense-reference distinction.

third party). As we have seen from the discussion of truth axiom (T), there is some confusion about whether epistemic logic is the study of attributions or claims of knowledge. Whereas in the latter we are dealing with speech and utterances, this is not the case in the former. Hintikka argues that some of the peculiarities of natural language are of a performatory character: they raise problems only when uttered. A textbook example of this is the Moore sentence: '$p \wedge \neg Bp$'. There is a fundamental difference between the following two utterances: $p \wedge \neg B_a p$, which can be read as: 'p is the case and agent a (not I) does not believe it', and $p \wedge \neg B_I p$, which can be read as: 'p is the case and I do not believe it'. Issues like these and the other mentioned issues are important examples of the many points in question that the modeller has to consider in her construal of knowledge.

First versus third-person perspective Another related issue to the previous topics of debate is whether knowledge is looked upon from a first-person perspective or from a third-person perspective. That is, if one looks at the situation from the viewpoint of the agent, taking into account the barricades and possibilities that the local epistemic environment of the agent may bring, or that one rather looks upon the agent from a God-like perspective, having complete knowledge of the whole situation. These perspectives are also referred to as the internal and (perfect) external perspectives. The importance of the difference between these different perspectives can be easily seen from the following example.

You (Y) and your brother (B) are in the living room of your parents' house. On the table is a mysterious box that might well contain chocolates. However, your parents forbid you to look into the box, so you both do not know whether there are chocolates in the box (p) or not ($\neg p$). When your brother leaves the room to go to the toilet, you secretly have a peek and see that the box indeed contains delicious chocolates. You carefully close it again, so that when your brother comes back, he is not aware of the fact that you had a peek and know whether there are chocolates in the box.

Now, from a third-person perspective, we have full knowledge of the situation. For example, from a third-person perspective it is known that there are chocolates in the box, that you know that and your brother doesn't, that he does not know that you know it, that you know that he doesn't know that you know, et cetera. From your point of view (first-person perspective 1), the knowledge about the situation is exactly the same as from the third-person perspective. However, from your brother's point of view (first-person perspective 2), the knowledge about the situation is different: he does not know that you had a look at what is inside the box, $\neg K_B(K_Y p \vee K_Y \neg p)$, nor does he know that there actually are chocolates in the box, $\neg K_B p$. That is, he has considerably less information than the overall information available in the example from a third-person point of view. This illustrates what differences there may be between a first and a third-person perspective, and that it is important to be aware of which perspective is taken. Models built from a first-person point of view may differ from models built from a third-person point of view.[3]

Whether a first or third-person perspective is taken depends on the field of science and the object of study. For instance, the first-person perspective is often taken in artificial intelligence and cognitive psychology, for studying the behaviour of autonomous agents.[4] On the other hand, in game theory, economics, and social psychology, one usually takes a third-person perspective. Although it has been argued by Hendricks (2005) that Hintikka initially meant to construe the logic of knowledge from a first-person perspective, it is clear that standard epistemic logic takes a third-person perspective. Epistemic logical models represent from an external point of view the way the situation is perceived by the agents. Furthermore, the modeller's information is perfect:

[3]But, as we have seen, the difference is not necessarily there.
[4]Suggested by Aucher (2010).

the modeller knows all the agent's epistemic attitudes and which world is the actual world (in case of a pointed model).

One may ask, as Hendricks does, why epistemic logicians usually take a third-person perspective, telling us which knowledge could possibly be obtained in a particular situation. After all, usually the agents do not have sufficient information to oversee the whole picture, which may be necessary for accessing that information. A possible answer given by Hendricks (2005, 33): "[o]ne recurs to an outside perspective for a principled answer that may then spill over into the local circumstances."

Guillaume Aucher (2010) studied the different perspectives that can be taken in epistemic logic, and gives recipes for switching between them. Although the first versus third person distinction is an interesting dichotomy that deserves further study, a more elaborate discussion of this will be outside the scope of this thesis.

Interconnectivity The topics of debate and issues discussed previously are all interconnected; sometimes very directly and strongly, at other times only weakly and indirect. The logician's anticipated aim of epistemic logic is connected to the extent to which she thinks epistemic logic is descriptive or normative. For instance, if one's aim is to improve our way of thinking (about knowledge), one views upon epistemic logic as normative. The connection between the type of knowledge that is studied and the aim of epistemic logic, just like the connection between the first or third-person approach taken and the aim of epistemic logic, is indirect, not necessary and not so strong. For example, one might envision the aim of epistemic logic to be understanding the epistemic concepts involved. To this aim it does not matter which type of knowledge is studied (ideally multiple types of knowledge are studied), or whether a first or third-person approach is taken (again, ideally both approaches are taken).

A citation by Lenzen nicely illustrates the relation between the debate on descriptivity versus normativity and the kind of knowledge that is studied:

> "Although [the epistemic-doxastic principles] ... are untenable as descriptive principles of what people actually believe and know, they must be considered to be analytically true for the idealised concept of implicit belief, implicit conviction, and implicit knowledge. And this very fact legitimated their use as epistemic logical principles." (Lenzen 1978, 64, emphasis erased)

That is, although epistemic logic cannot count as a descriptive account of actual, realistic knowledge as humans have it, it may count as an accurate description of an idealised version of it.[5]

Epistemic logic keeps rather neutral about the exact definition of knowledge, and focuses on the *logic* of knowledge. As Lenzen remarks:

> "The search for the correct analysis of knowledge, while certainly of extreme importance and interest for epistemology, seems not significantly to affect the object of epistemic logic, it is, the question of validity of certain epistemic-logical principle pause principles. The logic of knowledge is largely independent of whether knowledge is interpreted in the traditional sense of justified true belief, or in an any other stronger sense." (Lenzen 1978, 34)

Although in epistemic logic we might not be looking for a particular philosophical definition of knowledge, we could object that we still need to have an idea of what kind of knowledge we are studying. When making modelling decisions, one has to know what counts as knowledge.

[5] Cf. section 3.4.*ix*.

The interviews Having provided some background information on the topics of debate in epistemic logic and having attracted attention to their interconnectivity, we now have a look at how the interviewed researchers view upon the practice of epistemic logic. However, the views and opinions of different researchers diverge very much and sometimes even plainly contradict each other. Where someone says "Ultimately, of course, we want to model all the phenomena," someone else comments "I don't think that really the way people reason can be fully formalised." Epistemic logic has been called both "largely descriptive" and "clearly normative," and other researchers hold diverse positions in between these extremes. Also what to do with your findings from epistemic logic is debatable. Where one researcher states that "once you ... find out about some irrationalities in our reasoning, you should let people know and try to correct them," another researcher states the opposite: "We are not telling people how to reason about knowledge." Not only about the aims and perspectives of epistemic logic do the researchers disagree, also about its success criteria and function for the scientific community. Where one researcher finds that epistemic logic should lead to conclusions outside its own domain ("It's got to lead someplace outside, not just inside."), another researcher moots that "contributing to the internal research programme can be considered a success already." Likewise, researchers disagree about the role that epistemic logic plays for the broader scientific community. It has been stated that "epistemic logic turned out to be the common language that made it possible to bring together different professional communities," but this view has also been criticised to be "a bit colonial."[6]

As became clear from the interviews, researchers have very diverse opinions regarding the practice of epistemic logic. Their views on the aims, object of study, and methodology of epistemic logic diverge drastically. This is perhaps not all that strange: the work of different epistemic logicians may be of a completely different nature. For instance, work in epistemic logic can be very formal in nature, but it can also be much more philosophical, or even relate to research in social psychology. We need to distinguish between such different research agendas in order to make sense of the diverging and contradictory answers given by the respondents. In order to do so we introduce a categorisation, identifying four main research agendas of epistemic logic. We propose this categorisation based on our analysis of material from both the interviews (10 held for this thesis, 25 from the book *'Epistemic Logic: 5 Questions'*) and the epistemic logical literature.

In the following, whenever a reference to literature is not added to a citation, this citation stems from one of the interviews done as part of this thesis project. Furthermore, we use the abbreviation 'EL5Q' to refer to the book *'Epistemic Logic: 5 Questions'*, that can be found in the bibliography via (Demey et al. 2011).[7]

3.2 Defining several research agendas in epistemic logic

We propose to look upon the field of epistemic logic as consisting of four research agendas: the formal agenda, the philosophical agenda, the societal agenda and the computer science agenda. Before we elaborate on the definitions of these different research agendas of epistemic logic, we have a look at the choices underlying this categorisation.

The two dimensions underlying this categorisation are (a) the aim and (b) the object of study of epistemic logic. These dimensions are chosen to underlie our categorisation for two reasons. Firstly, they seem to naturally arise from the interviews. Researchers answered the questions in

[6]The citations in this paragraph have been the words of, consecutively: Williamson, Ramanujam, Romeijn, Leitgeb, Bílková, Aumann, Romeijn, Bonanno and Gierasimsczuk, in the interviews done for this thesis and the book *'Epistemic Logic: 5 Questions'*. The citations will be more are elaborately discussed in sections 3.3 and 3.4.

[7]We considered an abbreviation necessary, because the book is frequently referred to, and adding '(Demey et al. 2011)' every time after some other researchers' name would be confusing.

a certain way connected to the kind of research that they do. Sometimes this was even explicitly stated, like Halpern did: "I'm a computer scientist, so my goals are fairly pragmatic." Moreover, the different kinds of work described in the interviews naturally fit into the four research agendas that are formulated. Not only that, also did some researchers propose their own categorisations to account for different aspects of epistemic logic. These (partial) categorisations have been taken as a starting point for the one just presented. Secondly, for the current goal of discussing idealisations –and broader, the methodology– of epistemic logic, making the chosen dimensions (aim and object of study) underlie our categorisation proves to be very fruitful.

However, we should make sure not to take this categorisation to be more than it purports to be. For example, we should not consider the research agendas to be mutually exclusive and collectively exhaustive. What exactly this categorisation is and what it is not, and what role it plays in our study of the practice of epistemic logic, will be elucidated in the discussion at the end of this section. First, we have a close look at the proposed categorisation of the field of epistemic logic.

3.2.1 Formal agenda

"We are not really describing reasoning of real peoples. It's more like building logical systems which also take into account epistemic attitudes. ... Really I'm just building models of, lets say, human knowledge, but in this idealised way. ... I'm not sure that it's possible [to build models that really fit human knowledge], because I don't think that really the way people reason can be fully formalised."
- Marta Bílková

The formal agenda of epistemic logic comprises all technical and mathematical work on epistemic logical systems. Its aim is first and foremost to build good epistemic logical systems and proving these systems to have certain properties. Examples of research activities with a formal agenda are: proving certain systems to have certain metalogical properties (like soundness, completeness, decidability, and so on), proving certain theories to hold in certain systems, mathematically comparing different systems, et cetera.

The objects of study of the formal agenda of epistemic logic are logical or mathematical systems of knowledge.[8] Its claims are therefore conceptual claims about these systems, for example: 'this system is sound', 'in this system this particular theorem can be proven to be true', 'this system is a generalisation of this other system', et cetera.

Of course the motivation for formal work like this can come from outside the formal agenda, and usually it does. One might, for example, need to build a new logical system in order to be able to compare two philosophical arguments. However, the motivation for our intended application of formal work is not always this clear. It may also be that we formally study logical systems that interest us, without having a clear idea of what we can use them for. This will be discussed further in section 3.3.2 under the heading of 'Usefulness'.

3.2.2 Philosophical agenda

"The general attraction of epistemic logic has consisted in its capacity to give rigour and objectivity to the discussion of a philosophical question. Naturally, one must be careful not to lose sight of the philosophical question in one's search for rigour and objectivity!"
- Timothy Williamson (EL5Q, 252)

[8] Therefore, it falls largely outside of the traditional normative-descriptive distinction, as it is not (chiefly) in the business of representing (human) knowledge.

The philosophical agenda of epistemic logic comprises work that has strong ties to philosophical, in particular epistemological, debates. Its aims reach from *conceptual clarification* to substantiating and settling philosophical disputes. Examples of claims associated with this agenda could be ‘this formalisation shows an intuitive yet illuminating way of looking at knowledge’, ‘this formalisation proves that this example is not a counterexample after all’ and ‘this formalisation shows that this philosophical argument is inconsistent’. Moreover, the philosophical agenda has a certain substantive normative part in which claims are made in the form of ‘this is the way we should reason in order to get knowledge’.

The object of study of the philosophical agenda is (the philosophical concept of) human knowledge.

Next to conceptual clarification, there is also a normative part to the philosophical agenda of epistemic logic, which can be tied to the concept of *rational reconstruction*.[9] This means: taking a concept or phenomenon that is already there, and trying to make it more rational. This can mean, for instance, clarifying it, making it precise, making it more systematic, making it more fruitful or powerful, pointing to falsities and correcting them, or simplifying the concept that one started with. Rational reconstruction is an open-ended and pluralistic process that aims to give a better conceptual reality. According to Hannes Leitgeb, philosophy is rational reconstruction, and the philosophical agenda of epistemic logic in particular as well. He states:

> “For me the aim of [the philosophical agenda of] epistemic logic (and the semantics of epistemic logic) is on the one hand to rationally reconstruct belief and knowledge, and related notions like rationality or justification, but also to rationally reconstruct expressions standing for these phenomena.”

Leitgeb stresses that one always rationally reconstructs for a certain purpose, in a certain respect. For example, if we look at the notion of ‘belief’, we can rationally reconstruct this in different ways. We can rationally reconstruct it on a categorical scale, then we may get to possible worlds semantics. We can rationally reconstruct it on an ordinal scale, getting to AGM Belief Revision, for instance. Rationally reconstructing belief on a properly numerical scale, may get us to subjective rationality.

3.2.3 Societal agenda

“We use a lot of the tools that we use in dynamic epistemic logic to model what used to be something that psychology had been looking at, but now we can get a grip on, namely social epistemic and social psychological phenomena and informational lagoons and pathologies, like bystander effects, lemming effects, bandwagon effects et cetera.”

- Vincent Hendricks

The societal agenda of epistemic logic comprises the work that uses epistemic logic to describe and understand (social) epistemic behaviour. Its aims are largely descriptive, but may have a normative side to it when, for example, discovering pitfalls that can be avoided. The object of study in the societal agenda of epistemic logic is epistemic behaviour by individual or groups of real agents: people, groups of people, animals, et cetera. Claims in the societal agenda of epistemic logic may be, for example, ‘this system reflects how information transfer between groups of scientists works’, and ‘this formalisation shows how we can avoid this particular cascade’.

Epistemic logical work with a societal agenda is inspired by real-life phenomena, either observed by the researcher or described by researchers of another field. Often an alliance is sought with

[9] As suggested by Leitgeb in the forthcoming paper *‘Philosophy as Rational Reconstruction’* (2017).

researchers from these other disciplines, for instance with social psychologists, neuroscientists or economists.

3.2.4 Computer science agenda

"[Epistemic logic] is useful to be able to prove properties of computer programs, it is useful to conceptualise the way we think about security properties, notions like common knowledge are useful to understand things like coordination, it is useful to understand what properties the block chain has..."
- Joe Halpern

The computer science agenda of epistemic logic comprises the work in epistemic logic that has the strongest ties to computer science and artificial intelligence. These ties came into existence because epistemic logic proved to be useful in research on interpreted systems, and in particular, distributed computing. In this area of research, distributed systems are looked upon as communities of knowers: agents that exchange information and interact with the world. As Stalnaker stated, multiple-agent epistemic logic, "provided a precise representation, at the right level of abstraction, of the flow of information through the system, and of the character of various protocols governing such systems." (EL5Q, 243)

The aims of the computer science agenda of epistemic logic are fairly pragmatic, and closely tied to its intended applications in, for example, distributed computing. Other examples of the computer science agenda of epistemic logic comprise applications in artificial intelligence, like planning and designing intelligent multi-agent systems. Examples of claims could be 'this system is useful to prove a certain property of this particular computer program', 'this system is useful to conceptualise the way we think about security properties' and 'with this system I have verified this particular program'.

3.2.5 Discussion

As mentioned before, the proposed categorisation is not meant to identify mutually exclusive and collectively exhaustive research agendas in epistemic logic. Researchers can and will have research projects 'crossing borders': a project may, for instance, aim to describe some social phenomenon and in order for this phenomenon to be modelled some appropriate formal system may have to be built. In this case, the research project has (at least) a partially societal and partially formal research agenda. Often in the practice of epistemic logic projects will be covered by more than one research agenda. Consequently, we can also not tie individual researchers to a particular research agenda as just described. Furthermore, researchers may also change the kind of work they do over time. Just as researchers are alive and their work ever-changing, so is the field of epistemic logic a living field.[10] This field of science, as any field of science, is made up of its history and practitioners, and changes over time. We cannot, nor do we want to, pin down exactly what the field is or is supposed to do. Rather, we come up with some conceptual framework in order to be able to talk about the field of epistemic logic and its practice. This way we are able to escape the knockdown conclusions like 'everybody thinks differently about this' or 'it is very complex and depends on the context'. We can live with the vagueness and contradictions of views held among epistemic logicians, and still say something meaningful about them.

The presented categorisation is not based on a quantitative analysis, and it does not want to make any claims of this sort. Naturally, the current categorisation is just one among many possible categorisations of the field of epistemic logic, and one can think of many other distinctions underlying a categorisation. However, this categorisation provides us with a starting point upon which we can have a meaningful discussion of the widely diverging views that are held among

[10] I thank Johan van Benthem for attracting attention to this issue in one of our conversations.

epistemic logicians.

The terrain of *philosophy of epistemic logic* is new and largely untouched. Therefore, every structure that is being laid down in this terrain is new and up for debate. Although the presented categorisation can be seen to work well with respect to the current research aims (to bring order to the diverging views of epistemic logicians, to investigate the extent to which idealisations constitute a problem towards the practice of epistemic logic), other suggestions are welcome and likely to stimulate discussions of the subject matter.

3.3 Viewpoints regarding the practice of epistemic logic

Now that we have introduced a useful categorisation of the field of epistemic logic, we can turn to a meaningful discussion of the answers that were given in the interviews. For an overview of what was asked and why, and who was asked and why, we refer back to the methodology discussed in chapter 1.

3.3.1 Claims

Most of the respondents to the interviews agreed that claims in epistemic logic are mostly of a conceptual nature. However, as Nina Gierasimczuk and others remarked, there seems to be a growing need of empirical work.

What kind of empirical claims, then, might epistemic logic make? Van Benthem suggests that an empirical claim about some epistemic logical model could be of the form: 'with this model I said something interesting about reality'. This is still a relatively humble claim, reflecting that the model is only partially similar to the phenomenon it models. Hendricks is a bit more outspoken with regards to this question. He contends that in the earlier stages of epistemic logic, the models were "too crude," so that we could not do much else than conceptual analysis. This has changed nowadays, so that these models that started out conceptually now can be aligned by empirical material or be empirically verified. Hendricks describes this as "a long-running interaction between conceptual models on one hand, and empirical testing alignment on the other." For Leitgeb, when he speaks about rational reconstruction, which is covered by the philosophical agenda of epistemic logic, these empirical similarity adjustments are done just to remain similar enough to what you reconstruct. However, according to Leitgeb this is not one of the aims, as seems to be the case for Hendricks.

Rohit Parikh raises the problem of interpreting empirical data: they do not always send a clear message, so they are hard to work with. But this is only a problem with certain research agendas in epistemic logic. For example, you might mainly do formal work, and not be making any claims other than about the logical systems you work with. If this is the case, then problems with interpreting empirical data are not so much a problem for your research in epistemic logic.

Concluding we may say that, although claims in epistemic logic are still mostly of a conceptual nature, there has been a shift lately towards more empirically orientated research. Whether making empirical claims in epistemic logic is feasible is debatable and depends on your research agenda.

3.3.2 Success criteria

With the goal of finding out when and to what extent idealisations constitute a problem towards the practice of epistemic logic, one of the questions that were asked to the researchers was: 'what kind of considerations play a role in determining whether or not some model (or theory) is a success?' Of course, this depends on what the model was intended to do, and it depends on who you ask. Gierasimsczuk rightfully asked: "Who is to judge?" This is an interesting question that deserves to be studied in future research. For now, even without specifying who is to judge, many interesting answers were raised.

Adequate description

Those researchers whose goal it is to adequately describe knowledge as 'humans have it', consider getting as close of possible to this real-world phenomenon to be a measuring stick for success. This is mainly the case for researchers with a societal research agenda. For other aims, for instance in the philosophical agenda of epistemic logic, it may still be important to stay reasonably close to human knowledge in the real world. For example, if one tries to give a norm of knowledge or reasoning, it may not be a very useful norm if it deviates so much from human reality that it is impossible to even see how to live up to it or approximate it.

Mathematical beauty

As a criterion of success, amongst other things, Van Benthem mentions 'mathematical beauty'. This aesthetic quality is hard to grasp, but Gierasimsczuk puts this into words as follows:

> "in general, building logical systems is like building little castles which are supposed to hold together; beautiful structures that should enjoy these mathematically beautiful properties like soundness, completeness, neat computational complexity. Every such system is a success in its own right. It has to do with beauty, with mathematical elegance and originality, which is very difficult to capture."

When asked to what extent metalogical considerations influence the success of a model or theory, the answers are diverse. For some they can be considered to be a nice bonus (Van Benthem) and not crucial (Parikh), for others they can be "defeasible evidence that we have come up with a good rational reconstruction" (Leitgeb), or be useful for certain things we want to do with our system (Halpern). This connects us to the success criterion of usefulness, which will be discussed in the next paragraph.

Usefulness

For Baltag, the success criteria for an epistemic logical model may be a combination of good conceptual mathematical behaviour and expressivity. Both of these considerations are for him connected to usefulness. Halpern sees usefulness as a main criterion for determining success too, and for him metalogical properties matter as long as they enhance the usefulness of the system. The connection between metalogical properties like soundness, completeness, decidability et cetera and use are illustrated by Marta Bílková's words:

> "When I'm building a logical formal system to capture something, then maybe I want to use the system actually later. So I wouldn't stop at creating something that fits, like a glove that fits, but I would also wear it later. So I better make it wearable."

But this immediately raises a question that is not always easy to answer. Bílková continues:

> "And now the big question is: what are we using the systems for? I'm not sure I have the answer to that."

For some agendas, like the computer science agenda, it may be easy to pin down what your systems are used for. For the formal agenda of epistemic logic, this is not always the case. Of course, the systems are built so that they can be used, but what they will be used for is not always clear at the moment of construction. Baltag phrases it like this: "there is this implicit assumption that the logic that we investigate is useful for something. Because otherwise, why would you do it?" Perhaps we can compare the formal agenda to pure mathematics[11] here: also in pure mathematics researchers often do not have immediate applications in mind when they develop their systems.

Remarks of others can be seen as slightly critical towards such an attitude. Krister Segerberger remarks: "To be useful, models must be applied." (EL5Q, 204) Indeed, Robert Aumann even states that "the test of any important scientific theory or development [is that it also leads to conclusions outside their own domain.] It's got a lead someplace outside, not just inside." (EL5Q, 29) That not everybody agrees with this, can be seen from the penultimate paragraph in this section.

Leading to new insights

Halpern and van Benthem agree that when a model or theory leads to new (conceptual) insights, this can be considered to be a success. Van Benthem adds that unifying power may be a criterion of success too. As an example he mentions the iterated knowledge operators in epistemic logic that connect epistemic logical puzzles and the theory of mind with the concept of common knowledge in game theory.

Improve society and rationality

Parikh takes a fierce stance regarding the success criteria of epistemic logic: "I think [work in epistemic logic] is a success when it is integrated into political and social theory"... and: "epistemic logic should improve society." Somewhat along the same lines, Leitgeb argues that for the philosophical agenda of epistemic logic to be a success means to be actually getting more rational. This better way of looking at certain concepts, like knowledge, should also be communicated to the public, so that people can become more rational.

Community is interested

Mentioned by few researchers, but undoubtedly important as a success criterion, is that a number of people in your research community are interested in your work. Parikh mentions publishing papers and getting grants as (less important) aiming points. Van Benthem raises the topic of scientific progress, which is assessed by your colleagues of the research community.

Benedikt Löwe warns us for the danger of getting into an ivory tower situation:

> "We are driven by the fact that there is a sufficient number of people around us who consider this interesting enough that we continue. This is happening independent of having concrete applications. ... If you develop an area just by the fact that everyone around you tells you how interesting this is, then you will develop into an ivory tower situation. So you need to ground this every once in a while with things that connect you to the rest of the research world."

The danger of the ivory tower situation can be more severe for certain research agendas, for example the formal agenda of epistemic logic, than for others. The reason for this is, when doing research of this formal kind, one is not paying constant attention to particular applications that the research is intended to have. Compare this with the societal research agenda of epistemic logic, in which aligning your models with empirical data and research from other fields has a

[11] As opposed to applied mathematics.

much more central role. If there is the possibility of an outside check, whether it be to data or to the opinions of researchers from other fields, there is less danger of getting into an ivory tower situation.

All approaches valuable

Contrary to Aumann ("It's got to lead someplace outside, not just inside."), Jan-Willem Romeijn finds that contributing to the internal research programme of logic already makes a work in epistemic logic a success. Similarly, although Gierasimsczuk stressed that mathematical beauty and usefulness are important for epistemic logic, she also attracts attention to the strength of purely conceptual analysis. She does not want to think about any work of research in terms of failure (unless it is inconsistent); all approaches may contribute something interesting. Löwe, in same vein, argues that all approaches have the right to be utilised, and it is good that there are multiple approaches.[12]

Although all approaches can be said to have the right to be around, sometimes formal systems that have been put forward do not end up getting used by others, so they disappear with time. Baltag suggests that this may at times have to do with these systems being too expressive, because then they are not easy to use, consequently they will not be used much. Other problematic features of a system may be inconsistencies, or unintended behaviour of the system. In this case, the system may be either adapted, or in the worst case abandoned. Furthermore, Hendricks advances that, if in a certain part of epistemic logic there has been so much progress that it turned into an "industry of doing little tricks," we better drop it and move on. The reason for this being that it does not bring us much news.

Not one unique success story

Having listed all these possible views on criteria of success for epistemic logic, we should add that there is not one unique success story. Often it is a combination of different success criteria, a combination that depends on the goal and context. As has been often seen in the history of science, scientific models or theories can be first frowned upon and later glorified. Because the research community has an important say in which theories or models are a success or not, and the research community changes over time, the success of some particular work may change over time. Most researchers, when asked, gave a multifaceted picture of success criteria, a nuance that may have got lost in listing the different important success criteria above. To clear this up, let us end this section with Baltag's words:

> "The criteria of success are not unique. It's a trade-off between different criteria, some connected to reality, some connected to metaproperties. There is not one clear unique success story. ... It really depends on the situation if the same formalism is successful or not."

3.3.3 Intuitions

"Something's having intuitive content ... is very heavy evidence in favor of anything."
- Saul Kripke (1980, 42)

Epistemic logic is based on and connected to our everyday conception of knowledge. But what role should and do our intuitions about the everyday conception of knowledge really play in epistemic logic? Just like there is some variety in opinions regarding whether and how to use empirical data in epistemic logical research, there is quite some variety in what role intuitions may and should

[12] Cf. sections 4.2.4 and 5.4.

play. Although many researchers agree that intuitions can be very useful in guiding our research, conceptions diverge as to where the role of intuitions end.

Where some say: 'I have enough of an intuitive understanding of the concept of knowledge to do my research', others point to possible limitations to the use of intuitions. Gierasimsczuk raises the issue that intuitions are often very vague, and may have "little value for science." Bílková shares her worries: "intuitive understanding of how knowledge works ... can be very inspiring, but on the other hand there is this danger that intuitions are really subjective, and also may be not quite fitting." Therefore, she contends that we should always check our intuitions against (empirical) data.

Furthermore, multiple interviewed researchers note that intuitions (about the concept of knowledge) diverge, and that there is not just one notion of 'knowledge'. Different social groups may have a different conception of knowledge. Moreover, the feeling that I know something is not necessarily the same as knowing it. So how could the knowledge of one researcher (or the feeling that they know) be a trustworthy guide?

Because of reasons like these, many of the questioned epistemic logicians see only a limited role for intuitions. Gierasimsczuk, for instance, suggest that intuitions can only be used for questioning and discussing, not for judging science to be either good or bad. Also Hintikka sees a limited role for the use of intuitions. In his book *'Inquiry as Inquiry: A Logic of Scientific Discovery'*, Hintikka discusses the problem of generalising intuitions, as well as the limited access we have to them. This makes the extent of their applicability unclear, and it causes Hintikka to conclude that they may not serve as premises for (philosophical) arguments. Van Benthem claims that intuitions can be useful for "mapping out an area and putting constraints for an analysis," but can always be revised and refined. To him this is like music: we start with some musical sense, but by studying music our musical sense may change and be refined. Similarly, intuitions can only have the first word, not the last. Leitgeb agrees: for him intuitions can be prima face reasons to go for one formalisation (rational reconstruction) rather than for another, but we can always go back and revise these intuitions.

Hendricks argues that in epistemic logic intuitions can *only* be used as an illustration of outcomes of some formal framework. It cannot be used as more than this, in particular, we cannot base our formal framework on these intuitions unless it is reflected explicitly in the definitions, lemmata and assumptions of our framework. This can be contrasted to the practice of (mainstream) epistemologists, who "often remain quite vague about the tacit assumptions and presuppositions of their conclusions, which are based on intuitions and folksy examples." (Hendricks 2005, 16)

Another critique of the philosophical practice of this kind comes with Dennett and Hofstadter, who discuss the limited applicability of intuitions, especially when philophers' imaginations take flight:

> "When philosophical fantasies become too outlandish – involving time machines, say, or duplicate universes or infinitely powerful deceiving demons – we may wisely decline to conclude anything from them. Our conviction that we understand the issues involved may be unreliable, an illusion produced by the vividness of the fantasy." (Dennett and Hofstadter 1982, 230, cited by Hendricks 2005, 154)

Also Yoav Shoham (EL5Q, 228) underscribes a limited role for the use of intuitions, but then as necessary but insufficient conditions for a theory. He is, in fact, quite positive about their use: "One should be explicit about the intended use of the theory, and within the scope of this intended use one should require that every day intuition about the natural concepts be a useful guide in thinking about their formal counterparts."

Although intuitions can be very instructive in guiding our research, intuitions about the conception of knowledge are often prescribed a limited role. Of course we should not fail to distinguish two different objects of our intuitions here: our intuitions about how a proof might go versus our intuitions about the meaning of the word 'knowledge'. Using the first kind of intuitions is usually uncontroversially accepted, whereas using the second kind of intuitions is more debated and restricted. We have seen that the objections raised may imply that intuitions cannot be the judge of science or have the last word, should not serve as premises for arguments or as sufficient conditions for a theory, and that intuitions are not allowed to play more of a role than just for illustration of the formal framework outcomes. In the formal and computer science agenda of epistemic logic one makes use mainly of the intuitions of the first kind, and consequently there is not so much of a danger of using intuitions in the wrong way: using misguided intuitions in guiding you in proving a certain theory or building a certain system may merely cause you to take longer for the proof construction. This is in contrast with using the second kind of intuitions as, for example, a reason for dismissing or accepting a framework. The latter can be expected to happen mostly in the philosophical or societal agenda, as they tend to focus more on the human conception of knowledge. Therefore, as suggested by the researchers, in the philosophical and societal agenda of epistemic logic, the use of intuitions of the second kind should be restricted. At the same time the possibility of an outside check to our intuitions safeguards us from the subjectivity of these intuitions. This outside check is rather to be expected in the societal agenda, in which data about the real-world phenomenon of knowledge play are more vital role than in the philosophical agenda.[13]

3.3.4 How epistemic logic can be useful

Most researchers are of the opinion that the epistemic logical language, models, and structures are useful for something (outside the field of epistemic logic itself). There are, however, multiple viewpoints possible regarding in what way epistemic logic can be of use. Recall the discussion of the benefits of formalisation in the previous chapter. In section 2.2, the advantages of taking a formal approach were discussed along the lines of clarification, verification and exploration. In this section we will have a more elaborate look at what the interviewed researchers think that epistemic logic could be useful for, in what way and in virtue of what.

Van Benthem points to Popper's concept of the searchlight theory. If one wants to conduct research on real-world phenomena, one cannot just start by gathering a big amount of data. The data thus gathered would not give us any information about the real-world phenomenon, as we would have no framework to fit it into. Providing this framework is the duty of the searchlight theory. Van Benthem is of the opinion that epistemic logic can play this role, for example in experimental game theory or research on the theory of mind. Romeijn adds that epistemic logic could be useful for the forming of theories in such domains, "because it provides conceptual frameworks for understanding information transfer." Somewhat along the same lines as Van Benthem and Romeijn, Gierasimsczuk remarks:

> "I see philosophy in general, and particularly epistemology as a 'frontier', where we are allowed to touch issues that are not yet well-formed empirical questions. In order to conduct empirical research, we need to have a theory that allows us to state hypotheses or design experiments."

We may conclude that one of the possible ways in which epistemic logic can be useful is by aiding empirical research in the stages of forming research questions, interpreting data and developing

[13]But also outside checks may not be unproblematic, as is argued in sections 3.3.1, 5.1.3 and 5.2.1.

theories.

More in general, epistemic logic is also often looked upon as providing a new way of looking at things in other areas. For example, Ramaswamy Ramanujam suggests that epistemic logic can offer an interesting new way of looking at mathematical structures, and may raise new questions for the mathematical community. (EL5Q, 193) The relevance of epistemic logic is also often mentioned in connection to game theory. Artemov suggests that epistemic logic can offer us interesting new concepts for mainstream game theory, and form the foundation for the new discipline of epistemic game theory. (EL5Q, 12-5) Epistemic logic is also said to extend the conceptual apparatus of philosophy (Hintikka 1962, 10), and to work as "a catalyst for broader intellectual development." (Van Benthem in: EL5Q, 36) Summarising this, we may say that epistemic logic may serve the broader scientific community by providing a new way of looking at things, raising new questions and concepts, and creating interesting new research fields.

Epistemic logic may have a lot to offer for the development of the science of knowledge. For instance, as already discussed in the previous section, Gierasimsczuk indicates that epistemic logic, because of its idealisations, can point clearly to limitations of the human mind, thereby improving our understanding of human knowledge. Along the same lines, Hendricks sees a function for epistemic logic in identifying pitfalls in epistemic situations. By formalising certain possible real-life situations, epistemic logic may improve the way we look at or do things.

Now where in our human lives could this have an application? Romeijn attracts attention to situations in which knowledge claims serve as a basis for action. For example, this can be in court where judges do or do not convict the suspects, or in applications for grants by scientists. However, Romeijn is of the opinion that our current work in epistemic logic is often way too far from everyday situations of knowledge claims. The connections to the practice of knowledge claims should be made. Likewise, Hendricks sees a role for epistemic logic in respect to action based on knowledge (claims):

> "Knowledge is very good for deliberation, decision and action, individually and collectively. And that's the reason why you need an apparatus like epistemic logic, dynamic epistemic logic and other derivatives, to exactly pin down what sort of basis we are going to deliberately decide to act on individually and collectively, and what sort of pitfall do we find in the way."

Summarising this, we may say that epistemic logic could be very useful for society, because our actions are often based on knowledge or knowledge claims.[14]

Compare "I think that the notion of information cannot be usefully analysed without a matching logic," (Van Benthem in: EL5Q, 41) and "Philosophical discussions about the notions of knowledge and beliefs are interesting, but they dont lead often to somewhere if they are not formalised." (Williamson in: EL5Q, 150) In these two citations, epistemic logic is considered to be a formal tool that can be used to clarify philosophical arguments and enhance conceptual understanding. More generally, epistemic logic is often talked about in terms of being a 'toolkit'. We will discuss both what this toolkit can be used for and how it functions.

One of the merits of epistemic logic is that it is very precise about the assumptions that it makes, so that, for instance, tacit conditions may be identified so that one is clear about which method of

[14] However, Fitting still sees a gap in epistemic logical research here: "Generally our interest in the place of a person comes from our understanding of what those beliefs will cause a person to do. This seems to be missing from the standard model that is based on possible worlds." He suggests that we might be able to substitute possible worlds semantics by action-based epistemic semantics, to fill this gap. (EL5Q, 87)

reasoning is used in some model. In general, using a formal tool makes one aware of the modelling choices, which Timothy Williamson sees as a vital illustration of the usefulness of epistemic logic: "The role of epistemic logic, and that of logic, or formalisation in general, is that it helps us to become aware of what specific choices in the modelling of the concept lead us to." More specifically, Fitting (EL5Q, 89) discusses that in the process of modelling, "formal epistemic methods help us clarify what is basic to fields, or rather to our knowledge of and interaction with a field. In particular, which are the fundamental concepts, and which are the derivative ones." A good example of this would be the interesting new insight at the birth of the field of epistemic logic of taking the concept of ignorance as fundamental in researching knowledge. Since then knowledge of some proposition is traditionally defined as having excluded all the possible worlds contradicting that proposition. Some other examples of where the precise assumptions of epistemic logicians may be helpful are: testing coherence of different philosophical approaches (Williamson), disentangling puzzles (Arló-Costa) and identifying necessary conditions for coordinated actions (Shoham).

Van Benthem remarks that "the models used are extremely simple, but that is precisely why they allow us to see essential structures, and develop major themes in a perspicuous manner." (EL5Q, 37) Van Benthem is one of many that see merit in the simplicity of epistemic logical models. When asked when his interest for epistemic logic came about, Stalnaker answered that it "provided a precise representation, *at the right level of abstraction*." (EL5Q, 245, emphasis added) By eliminating some complexity of epistemic issues, epistemic logic can be very helpful for several purposes. It has been mentioned that the simplicity and systematicity of epistemic logic could help reveal deeper relations between different epistemic concepts (Arló-Costa), and bring to light subtle points (Bonanno). Horacio Arló-Costa also contends that "some deeper problems in the theory of knowledge and belief, concerning the relations of knowledge with truth and self-reference and belief with degrees of belief are only evident when one has the proper logical tools at one's disposal." (EL5Q, 4) Both the simplicity and the systematicity or rigour of epistemic logic are often praised for their illuminating properties. Artemov puts it like this: "Formal methods do not usually resolve fundamental philosophical issues but they provide the necessary reliable framework for more advanced reasoning about these issues." (EL5Q, 17) Summarising, we might say that epistemic logic could aid in our understanding of certain concepts and problems by being both very simple as well as systematic. Besides this, we might add relative 'objectivity' as an attractive feature of epistemic logic for formalising philosophical disputes. (Point raised by Williamson, (EL5Q, 252))

Hendricks explains how epistemic logic can aid in conceptual understanding, because of its local approach:

> "Instead of pursuing a global conceptual understanding of knowledge, formal epistemologies proceed in a more piecemeal fashion. A certain amount of conceptual understanding is presupposed, or a certain set of conceptual parameters are fixed. ... Although the results of the piecemeal conceptual analyses in a formal framework may not add up to a global concept of knowledge,[15] they may all the same reveal something about the structure of the ingredients making up such a concept." (Hendricks 2005, 15-6)

Besides a useful 'toolkit', epistemic logic has also been looked upon as a 'common language' in which discussions of epistemic issues can be phrased. Ramanujam spoke about epistemic logic in terms of "a very attractive way of discussing things." (EL5Q, 191) Melvin Fitting and Giacomo Bonanno agree that epistemic logic may enable talking and reasoning about epistemic concepts like knowledge and belief. (EL5Q, 51, 89) Along the same lines, Aviad Heifetz remarked that

[15] However, he also argues that these local approaches may contribute to global understanding of the subject matter. (Hendricks 2005, 162)

“as a language, epistemic logic provides a vocabulary to communicate and express ideas about interaction and mutual knowledge and belief between individuals.” (EL5Q, 139) More outspoken, Bonanno takes epistemic logic to be “the common language that made it possible to bring together different professional communities.” (EL5Q, 48) Not everybody agrees. Gierasimsczuk criticises the view of epistemic logic as the common language: “The idea that logicians will be able to sell the language that they designed as the common multi-disciplinary language is a bit colonial, I would say.” Summarising this, we may say that epistemic logic is sometimes viewed upon as a helpful language that may simplify discussions of epistemic issues, but regarding it as ‘the common language’ of multidisciplinary knowledge-related research may for some be a bit too strong.

3.4 Idealisations

Having had an elaborate look both at the aims and methodology of epistemic logic, we may now turn to a discussion of different possible ways of dealing with idealisations in epistemic logic. To this end, we start with looking at Lenzen’s words on this:

> “[If one looks at the actual knowledge of human beings,] then even most elementary ‘laws’ ... would not be valid, and one could hardly find *any* epistemic logical law which adequately describes the factual knowledge- or belief-system of an arbitrary subject, a. However, this sceptical conclusion rests on a very narrow conception of our everyday’s attribution of propositional attitudes. ... In everyday’s discourse, however, we standardly presuppose that the people with which we talk have an adequate understanding of what is said. Therefore we assume that their belief- or knowledge-systems satisfy certain conditions of rationality, in particular a certain amount of logical consistency and deductive closure.” (Lenzen 2004, 965)

Lenzen’s reaction on the inherent idealisations of epistemic logic reveils a variety of possible ways to deal with this matter. Firstly, he observes that if one seeks to construe a logic of knowledge that represents human knowledge in a realistic way, this does not give us a very interesting logic to work with. Secondly, he remarks that this sceptical conclusion bears on a very narrow conception of knowledge; perhaps we should look for a different conception of knowledge. Thirdly, he argues that in real life we presuppose a certain degree of understanding and rationality of the people that we speak to. This is both an argument for saying that epistemic logic as we know it is sufficiently close to real-life situations (that is, the axioms of epistemic logic are acceptable, because in real-life situations we also presuppose a certain degree of rationality), and an argument for construing knowledge with respect to some rational agent.

As Hendricks and Symons put it: “Epistemic logic inevitably traffics in idealizations,” and there will probably be not one logician that denies this. However, as it can easily be seen from Lenzen’s words cited above: there is a variety of ways in which we can react to this given. For a start, we could take the idealisations as a reason for being sceptical about the whole field of epistemic logic, as many philosophers, among whom famously Max Hocutt in his paper *‘Is epistemic logic possible?’*, have done. Let us call this reaction: *(i) epistemic logic is doomed.* To avoid this sceptical conclusion, we may point to the fact that if we do not idealise, we do not get anywhere either. Better an idealised theory than no theory at all. Call this second type of reaction: *(ii) not idealising does not get us anywhere either.* A third way in which we could react to the idealisations inherently present in epistemic logic, is to argue that the model is sufficiently close to real-life reasoning about knowledge, and therefore admissible. Call this reaction: *(iii) sufficiently close.* Furthermore, many researchers regard the idealised model as a (suboptimal) starting point for our models that can be further refined later on. Call this: *(iv) idealisations as a starting point.*

In general, we may give a pragmatic justification of the idealisations: if the model is good enough for its intended application, why would you change it? Call this justification: *(v) pragmatic.* In many cases we can learn from idealised models: *(vi) we can learn from idealisations.* But also the objections that can be raised against certain idealisations have made researchers aware that we need to look for the right balance in our systems, the balance between having nice mathematics to work with and staying close to the real-world phenomenon we are trying to study. Call this reaction: *(vii) we should find a balance.* Another possible reaction is that our model does not represent the real world, but rather some idealised world. Now there are two ways in which we can look at this idealisation: either we say that our knower is some kind of idealised agent, or we say that the concept of knowledge that we are studying is idealised. The first approach we call *(viii) rational knower*, the second approach we call *(ix) implicit knowledge*[16]. The last possible reaction to the problem of idealisations that we consider is that we should be open about the idealisations we make and remain aware of them. We call this last reaction: *(x) we should be aware of the idealisations we make.*[17]

As can be seen from Lenzen's citation, these different reactions are often connected and intertwined. However, to create some order out of chaos, in the following we discuss all these types of reactions separately.

(i) epistemic logic is doomed

Although not many epistemic logicians will take this rather sceptical route, it is good to have a look at what critics might say about the practice of epistemic logic. A really easy way to attack and condemn epistemic logic is to maintain that the axioms, and the idealising assumptions that are made, are unrealistic and therefore undefendable. One might take the inherent idealisations of epistemic logic as a reason for saying that epistemic logic is a misguided practice. However, it is important to note that the force of such critiques depend on the intended aims of epistemic logic, and in general the context of the modelling situation. This issue will be discussed more elaborately in chapter 5.

Hocutt famously criticised epistemic logic as a misguided practice, and argued that "every example [of epistemic logic] faces the following dilemma: either it does not have anything especially to do with knowledge and is therefore epistemic in name only, or it does and is, in consequence, logic in name only." (Hocutt 1972, 433) Many others have argued that epistemic logic is unrealistic and therefore misguided. For example, Hector-Neri Castañeda argued in his review of Hintikka's pioneering work, that Hintikka's notions of knowledge and belief are "much too strong," because there are probably "no human instances" of it. (Castañeda 1964)

(ii) not idealising does not get us anywhere either

As Lemmon early on noted, "[a] realistic logic of knowing contains no distinctive theorem apart from [the axiom of truth] and its logical consequences." (Lemmon 1967, 78, cited by Lenzen 1978, 14) Avoiding the pessimistic conclusion that sceptics and Hocutt reach, one may object that not idealising does not get us anywhere either. A realistic logic of knowledge would not yield anything interesting. Furthermore, real human knowledge is a very vague and varying concept, both its rules and deviations of these rules are not systematic enough in real life and cannot be

[16]Note that the logician's idealised concept of 'implicit knowledge' (referring to available information) is to be distinguished from the concept of 'tacit knowledge' used by psychologists and neuroscientists, which is –confusingly enough– sometimes also referred to as 'implicit knowledge'.

[17]Of course we can imagine that there are more reactions possible, but these ten reactions seem to be the most pervasive in the literature and interviews.

realistically modelled by any logic.

"[T]he world's descriptive complexity is literally limitless," writes Rescher (2005, 101). Imagine that we would like to model human knowledge realistically, and take into account all possible contextual parameters that may play a role in the attribution of knowledge. It would amount to setting an enormous set of parameters. For example, if we would like to describe a particular person of average reasoning capabilities and rationality, and we set parameters for her (general) knowledge base, the number of thinking steps she can perform, memory recall of a certain degree, logical abilities, et cetera, we might infer that this person can make reasoning step A in situation B. However, if this person were, say, in love, we might not expect her to draw this basic inference. Consequently, we would need to also include a parameter describing her state of being, that is, whether or not she is in love, drunk, in a hurry, stressed out by her boss, and so on. No matter how detailed our set of parameters, there will always be more possible factors that are important for the attribution of knowledge in a certain situation. Besides, human behaviour is erratic and it would already be an idealisation to say that person x always performs three thinking steps, no more, no less. Similarly, Rescher (2005, 5) remarks that "reality, nature, has an effectively infinite cognitive depth in point of detail, in that no matter how elaborate our characterizations of the real, there is always more to be said.' However, he argues that the impossibility to grasp this diversity and infinity is not a reason for abandoning the practice of epistemic logic.

Also Hintikka acknowledges the "richness and fluidity of ordinary language", and does not take it as a reason for abolishing or adapting his theory: "recognising this multiplicity has not necessitated any changes in the relatively simple rules which are basic to our study, or diminished their importance." (Hintikka 1962, 123)

Not only in the literature, also in the interviews researchers make similar remarks. For example, Van Benthem notes that in lifting idealisations there is the danger that the theory might become too complex to work with, and therefore does no longer yield useful insights. Baltag phrases it as follows:

> "The problem is ... you cannot have a theory of real-life... The other thing is: it's better to have a theory than to not have one, because a theory helps anyway. Because between the idealisation, the simplified model, and the real situation there are some similarities. In certain contexts, in certain situations, your theory - no matter how idealised - will work: it will actually predict reality, so it will give you power. And also, when it is about normative style, it will give you a way to correct reality."

So, rather than mourning about the lacking realism of the assumptions of epistemic logic, we might accept the idealisations and have something to work with, to possibly learn something from it. This seems to be the conclusion of many researchers in both the literature and interviews about epistemic logic.

(iii) sufficiently close

As we noticed before, in his discussion of the idealisations of epistemic logic, Lenzen implicitly argues that epistemic logic, although it may not seem that way, is in fact quite close to everyday's discourse. After all, in real-life scenarios we also presuppose some form of rationality of the people whom we talk with. He also contends that "genuine counterexamples [to the epistemic idealising assumptions] are more rare than generally thought." (Lenzen 1978, 61)

A reaction in this spirit, hinting that knowledge in epistemic logic is not all that different from knowledge as we look at it in everyday life, we find more often in the literature. For example, Rescher (2005, 5) describes the connection as follows: "The reality of it is that epistemic logic is an applied logic and its uses, being geared to salient features of the established concept of knowledge, stand correlative to the ways in which we actually do talk and think about the matter."

Another reaction is trying to make epistemic logic approximate reality closer and closer. In various modern works several parameters are added to the picture, like awareness, relevance, levels of knowledge et cetera. In the next paragraph, we discuss taking ideal models as a starting point from which to get to more refined models.

(iv) idealisations as a starting point

It has often been suggested that idealised models may serve as a starting point, on the basis of which we may then construe more realistic models, or models that better suit our purposes. Similarly to what has been suggested by Baltag in subsection (ii), Hendricks notes:

> "you don't start out with the realistic model, because they are too many parameters to worry about. You start out with something idealised, and once you get that nailed down, you start looking at more contingent or additional phenomena that you might try to model."

However, the issue of taking idealisations as a starting point has two sides to it: it can be very fruitful to start with an extremely simplified model, but it may also come with a warning; that we first need to understand the simplified model before we may expand. The latter point has been raised recently by Van Benthem, who comments that we should first understand the "fruitful simplifications" that epistemic logical models make, before we can extend our research. (EL5Q, 44) Dana Scott, some decades earlier, already made a far more negative claim regarding this topic:

> "Formal methods should only be applied when the subject is ready for them, when conceptual clarification is sufficiently advanced... I feel that insufficient consideration has been given to questioning appropriateness of results... It is all too tempting to refine methods well beyond the level of applicability." (cited in EL5Q, 234)

This raises the question: are the models of modern epistemic logic refined beyond the level of applicability? This will be discussed further in the discussion and conclusion in chapter 5.

Be that as it may, the call for more realistic models is pervasive. Researchers from different areas call out for credible models of cognitive real play, models to better understand bounded reasoners, et cetera. Halpern thinks that many applications of epistemic logic could benefit from a realistic model of bounded reasoners: "To the extent that we are using epistemic logic to understand things, and the things that we are trying to understand are imperfect reasoners, it would be good to get a logic that can help capture that." However, as we may recall from the previous section, Halpern does not think that there are any really great methods for modelling resource-bounded agents so far.

For some applications, like cryptography, reasoning should be a vital part of the formal theory, and cannot be modelled by some formal system in which reasoning turns out to be superfluous. When we encrypt some important document, we do this under assumption that nobody (without our private key) has sufficient computational power to brute force[18] the decryption. A logically omniscient agent has no limits to her computational power, so for certain applications we just assume too much of the agents. That is the case for features that hold for all epistemic logical models (like logical omniscience), but also for features that come with a particular axiomatic system. Until the 90s, researchers debated about which system would best capture certain types of

[18]Breaking the code by trying all possibilities.

knowledge. At some point, this discussion settled down, according to Shoham way too early: "the discussion around them seemed to subside long before closure was reached, which makes it hard to build on those foundations as we try to apply them and/or extend the domain of discourse." (EL5Q, 225) Also Hans van Ditmarsch found it "surprising that the attention for alternatives for **S5** in the 70s and 80s has ground to a halt." (EL5Q, 77)

Although idealisations have served their goal in providing us with a starting point upon which to base our epistemic logical research, there are still many reasons why we should look for more realistic models. This could be by refining those models based on mathematical modal logic, or by looking for completely new mathematical structures that can be given an epistemic interpretation. Bílková urges us to be creative: "To resolve (some of) the idealisations, we should have an open mind set and use all mathematics that we have ever seen. Sometimes we need to start from scratch. We should rethink using modal logics."

(v) pragmatic

As Stalnaker notes,

> "It is interesting, and puzzling, that theories that make such egregiously unrealistic assumptions can nevertheless seem to be fruitful and clarifying. That does not mean we can ignore the problem, but I think it does give us reason to carry on, the hope that a solution or at least a clearer view of the problem, will emerge as the framework is developed and applied." (EL5Q, 246)

This more pragmatic view of evaluating idealisations in epistemic logic will be discussed in the current paragraph.

Lawrence Moss moots:
"the fact that the representations [of dynamic epistemic logic] mostly do work is a big point in its favour." (EL5Q, 171) Halpern agrees that for analyses that do not depend on resource-boundedness, "it's interesting to study it in this idealised way and it may give us useful results." He urges philosophers to take the pragmatic aspects of epistemic logical research into account more:

> "Philosophers are much too concerned with trying to construct examples to shoot down theories, and not concerned enough with the pragmatic aspects of questions like causality and epistemic logic. I think they would get a lot further if they would focus more on these pragmatic issues."

Hendricks always looks at the advantages and disadvantages when modelling epistemic behaviour:

> "You've got to look at the pros and cons of it. If you have common knowledge, then you can all of a sudden model trade and no trade, you can do agree to disagree, et cetera. The pros by far outnumber the cons, as far as I can say. And from that perspective we are pragmatists here! It's not like you have to take a stand: if it works good, then don't change it! ... If it doesn't work: change it, find something else that works."

Whether the pros outnumber the cons or not depends on our research aim and context. Stalnaker argues that the idealised assumptions of epistemic logic may well be helpful with respect to certain applications:"[idealised] conceptions of knowledge ... may not provide plausible analyses of knowledge generally, but ... may provide interesting models of knowledge that are appropriate for particular applications, and ... may eliminate, in an idealised way, one or another of the dimensions of the complexity of the terrain. (Stalnaker 2006, 170)

As Stalnaker shows, one can also point to useful applications of some system in defence of the epistemic axioms. Recall from the previous chapter the reasons that were given in the *'Handbook of Epistemic Logic'* for the acceptance of the **S5** properties: because this system is really useful for modelling certain computer processes (for instance, an interpreted system) the axioms are appropriate.

Concluding, we may say that, when evaluating idealisations in epistemic logic, we might take a pragmatic stance and see whether our model does what it is supposed to do. When looking at idealisations pragmatically, the intended aim and use of the model takes a central role. While a model may be idealised, we can still learn from it. The following paragraph will take a closer look at the ways in which we can learn from idealisations.

(vi) we can learn from idealisations

Even though a formal system might not always be all that similar to the target system[19], idealisations can teach us something on their own. For example, as Löwe suggests, they could teach us something about a relationship between the formalisation and the target system: "The idealisation going beyond what is the concrete problem is also important, because it tells us something about the relationship between the ideal mathematical world and the real world." By studying something outside what is actually possible, you may find out more about the actual.[20] Now what could we find out about actual human knowledge by looking at idealised models of knowledge? Gierasimsczuk suggests that idealised epistemic logical models help us understand the limitations of the human mind. Also Rescher recognises this:

> "while such an over-generous conception of knowledge construes this idea in a somewhat unrealistic way, nevertheless, it does have a potentially useful part to play in epistemic deliberations. For when it comes to inquiry into the *limits of knowledge*, this conception will prove serviceable. For if and when there are any facts that cannot be known in even this over-generous sense of the term, then such facts are thereby securely emplaced outside the realm of the knowable at large." (Rescher 2005, 15)

Not only can we learn from comparisons of the ideal to the real, Van Benthem defends that idealisations might broach a whole new area of "surprising new applications and research directions, far beyond their international initial territory and the intention of their originators." (EL5Q, 44)

Slightly less positive, Williamson suggests that we may turn the negative points that he and others have raised against using the modal system **S5** for epistemic issues into a more positive direction by asking ourselves how sensitive the results are to the assumption of the **S5**-axioms. Then we can generalise the more robust results by weakening the assumption of **S5** to a less strong modal system. This can teach us interesting things about the scope of applicability of the earlier achieved results to a broader class of modal systems.

Sometimes, however, making the idealising assumptions that we do, can plainly give us a better result than using a more realistic system. A great example of this was raised by Baltag: when the stakes are truly high, for example in a state of war, you would like to consider your agent to be logically omniscient and computationally unbounded. For if you assume, for instance, your opponent to have some particular computational limit, you might get the Enigma situation as

[19] Recall from section 2.2 that by 'target system' we mean to refer to some situation, concept or phenomenon in the (physical) reality.

[20] Cf. the biological study of three-sexed organisms, as discussed in the next chapter, section 4.1.4.

was seen in World War II.[21]

Summarising, we may say that there are various ways in which we can learn from idealisations: we can learn something about limits of knowledge because of the model's deviance from reality, we can learn despite idealisations by looking at robustness of results, we can learn from idealisations because for some applications idealised models are more appropriate than realistic models, and we can learn from idealisations in ways that we do not even know yet.

(vii) we should find a balance

Löwe remarks that in the process of epistemological modelling, we are looking for the right balance between being realistic and mathematically interesting:

> "There is this mathematical model on the one hand, which is extremely clean and gets precise answers. And on the other hand there is the concept that we are modelling, and that is much less clean and very complicated. We have to find the balance..." Sven Ove Hansson agrees that using logic in epistemology includes an idealisation process, which inherently "involves a trade-off between simplicity and faithfulness to the original." (EL5Q, 129)

Of course, where to find this balance depends on the research question and the context of the modelling situation. It is impossible for us to give a general recipe for the perfect balance between being too idealistic and too realistic for any modelling situation.[22]

(viii) rational knower

Sometimes researchers defend idealisations by stating that our object of study is not a real-world phenomenon, but rather some idealised version of it. There are two ways to go about: either our epistemic agent is idealised, or the concept of knowledge that we study is an idealised version of it.

> "Our results are not directly applicable to what is true or false in the actual world of ours. They tell us something definite about the truth and falsity of statements only in a world in which everybody follows the consequences of what he knows as far as they lead him. ... They are applicable to actual statements only insofar as our world approximates one of the 'most knowledgeable of possible worlds' ... The applicability of our results may thus be said to presuppose a certain amount of rationality and the people whose attitudes are being discussed depend on an assumption of rationality." (Hintikka 1962, 36-8)

When an epistemic logician talks about some idealised knower, the magical word 'rational' usually comes into the picture. Also Lemmon contends that "[w]e must view epistemic logic as giving the logical truths concerning a logical fiction, a sort of 'ideal knower', the rational man." (Lemmon, in Chisholm 1963, 773-95, cited by Hocutt 1972, 436)

What this rationality constitutes, then, is –somewhat circularly– defined in terms of the rationality axioms that happen to coincide exactly with the modal (or epistemic) axioms. It may be

[21] The Germans expected that the Allies could not break the code, as the Germans themselves did not have sufficient computational resources to do so. The British invested a lot in breaking the code, and their success changed the course of war dramatically.

[22] Likewise, in the physical sciences (discussed in the next chapter) we will find that what counts as relevant similarity for a formal system with regards to the target phenomenon always depends on the details at hand. The respects in and the degree to which the formal system should resemble the target system are determined by research question and context.

objected that just calling these agents rational does not paint the complete picture. An agent may be completely rational (in the intuitive sense of the word), and still be bounded in its resources, for instance, the agent may be rational while lacking the time for completing infinite inference steps. We should be careful to distinguish here between the philosophical or intuitive notion of rationality and a formal notion of convenience. Epistemic logicians often appeal to some kind of idealised agent that is postulated to be fully rational.

(ix) implicit knowledge

"Using modal logic to formalise knowledge and belief suggests that one has an idealised version of these notions in mind." (Van Ditmarsch et al. 2015, 46)

Sometimes in reaction to the problem of idealisations in epistemic logic, one tends to see the knowledge that is being studied as an ideal kind of knowledge, in contrast to a more realistic notion of knowledge as humans seem to have it. The idea here is that the knowledge being studied is not the actual knowledge of an agent, but possible knowledge that the agent might have had if she had sufficient resources and rationality. Recall from the discussion of system **K** in section 2.3.3 that Rescher objected that, if one accepts its axiomatisation, we are studying available information rather than knowledge proper. Oftentimes this concept is referred to as 'implicit' knowledge. This terminology is especially popular among computer scientists, that often construe knowledge as an **S5**-modality.

Hintikka was the first epistemic logician to take refuge to this kind of knowledge.[23] He argues that an agent might fail to be aware of some consequence of her knowledge, but if she were asked about it, that is, made aware of it, she would know. This makes the conception of knowledge an implicit kind of knowledge, the kind of knowledge that the agent *would or could have if she were asked about it* and made to ponder on it.

Lenzen, in his survey of epistemic logic of the sixties and the seventies, contends that "[t]he idealisation is neither unrealistic nor far-fetched. It only involves a commitment to affirm explicitly what has already been affirmed implicitly," (Lenzen 1978, 64) and thereby gives the start signal for an interpretation of knowledge as 'implicit' knowledge.

Rescher recently referred to the idealised concept of knowledge as 'available information':

> "However convenient this sort of 'knowledge' is from a logician's point of view, it is totally unrealistic as a representation of the sort of knowledge that figures in everyday language discourse about this idea. In fact, what is at issue here is not knowledge proper but rather available information." (Rescher 2005, 15)

Whatever we call it, the idea in all these approaches is the same: the concept of knowledge that is being studied in epistemic logic is not an everyday concept of knowledge, but rather some idealised concept of knowledge. Changing our object of study like this, we no longer expect the agent to be able to carry out all the inferences of her knowledge. Doing so, we are no longer talking about a realistic notion of knowledge.

[23] In his seminal work, he does so by defining the concepts of 'defensibility' and 'self-sustainability'. A statement is defensible if it is true in at least one for the agent accessible worlds, and self-sustainable if it is true in all her epistemic alternative worlds. (Hintikka 1963)

(x) we should be aware of the idealisations we make

A whole different kind of reaction to idealisations in epistemic logic is that idealisations may not constitute a problem towards the practice of epistemic logic, as long as we are open about which idealisations come with our system and remain aware of them. Löwe presents the issue as follows: "The problem that sometimes occurs with these formal systems, is [that] researchers are not clear about what it actually is that they are studying. If they were honest and said: 'this is a formal system...', then sometimes the misunderstandings would not occur that do occur." Also Bílková stresses this honesty: "we should be honest about what idealisations we make, we should make it very clear if we can... so that they are open for discussion."

These two remarks seem to hold for all research aims that one can have in epistemic logic. When the aim of our epistemic logical research is descriptive, however, we may not be there yet with only honesty. Leitgeb: "[when] I'm smoothing out [my model or theory], as long as I am aware that I'm doing that, that's also fine, but I should be aware that in principle I'm messing now with the truth, and there is something worrisome about that if you're in the business of science."

Discussion

As Baltag notes, "All science is based on idealisations, no matter what you study." However, there are different ways of reacting to these idealisations. The current section has given an overview of possible reactions in epistemic logic. This is by no means meant to be exhaustive, but it has touched upon the most pervasive opinions that are held regarding the subject matter.

As discussed, it is not possible to give a general account of when idealisations constitute a problem to the practice of epistemic logic and when not. This always depends on the context, as was also described by Shoham: "the logic proves useful to reason about certain aspects of distributed systems, and the mismatch between the properties of the model operator case and everyday word 'know' does not get in the way, within these confines. All this changes as one switches the context. For example, consider cryptographic protocols[24], then the K axiom is blatantly inappropriate." (EL5Q, 228)

As the role and acceptability of idealisations switches from context to context, it has to be decided in every single new situation whether or not certain idealisations constitute a problem towards the aim. Halpern: "I think it is up to the modeller to think hard about what is appropriate in his or her application." Although we may agree with this, we may still try to give the individual modeller some handles to assess the situation. In order to do so, we first have a look at the lessons we can learn from a discussion about abstractions and idealisations in the natural sciences.

[24] Recall the cryptography example discussed in section 3.4.*iv*

Chapter 4

Abstractions and Idealisations in Science

If we want to discuss abstractions and idealisations in epistemic logic, we would be wise to consult philosophers of science who have written about this subject. Even though they tend to focus on physical science, we may compare their views if we do not plainly overlook the differences in approach and between the research fields.[1]

Abstractions and idealisations are much discussed topics in the philosophy of science and have significantly gained in attention during the eighties. Different authors present different views and different names for similar but often slightly varying concepts. An analysis of their views is complicated by the fact that these authors often use identical words to refer to different concepts. In order to use their ideas, we need to have a clear idea about which concepts these authors refer to when using words like 'idealisation' and 'abstraction'.

In this chapter we purport to give an overview and a comparison of the views of leading authors in this field. This chapter treats just a small selection of those many authors who have concerned themselves with the topic of idealisations in science. However, the few different approaches and definitions outlined below give a good impression of the variety of usages present in the philosophy of science, and raise interesting issues concerning abstraction and idealisation. The chosen selection provides insights in the differences in the underpinnings and objectives of the various concepts. This will give us just the handles we need to continue our investigation of idealisations in epistemic logic.

Models

When speaking of 'idealisation' or 'abstraction', abbreviated to "representational imperfections[2]" by Martin Jones (2005, 176), one usually speaks of an idealisation or abstraction in relation to a model. One can also speak of representational imperfections as concern laws or theories, diagrams, graphs et cetera. The selected authors all speak of representational imperfection in connection to models, Nancy Cartwright additionally speaks about laws and theories. With the purpose of this thesis in mind –discussing idealisations in models used in epistemic logic– we will focus on models only.

Ernan McMullin warns for the different uses of the word 'model'. (McMullin 1985, 257) Whereas a mathematical logician commonly uses it to describe a set of sentences that makes a certain set

[1]This will be further discussed in section 5.1.

[2]Recal that 'representational imperfections' is intended to be a neutral word that includes all types of deviations from the reality of the target system.

of propositions true, the architect usually reserves it to talk about a scale model: a detailed but often smaller physical model of some construction. To achieve a more thorough understanding of the connotations of the word 'model' as well as the discussions involved around modelling, models and representation, we start with a glance at the various stances occupied by philosophers of science.

4.1 Models in science

"Scientists, whose business is understanding the empirical world, often spend their time considering things that are known not to be part of that world." (Godfrey-Smith 2009, 1)

Frictionless planes in physics, ideal gases in chemistry, games of perfect information in economics, dynamics of infinite populations in biology, situations with logically omniscient agents in formal epistemology... Models, and often highly idealised, unrealistic models, are omnipresent in today's sciences; they are one of the principal instruments of modern science.

4.1.1 Brief history of modelling

Although models are ubiquitous in modern day science, they have been in use since the early days. Galileo's diagrams may serve as a textbook example of a model. In his writings Galileo proposes to represent problems in physics geometrically, so that a solution to the problem can be obtained by mathematical means. So doing, Galileo *changes the question*[3], and after a geometrical solution has been found, the result has to be translated back in terms of the physical reality again. This step of changing the question is central to the practice of modelling.

In the literature philosophers of science often make it seem like modelling is the only way of theorising. Nonetheless, it has to be said that modelling, though perhaps nowadays the most important one, is just one among various strategies of theorising. An example of another strategy of theorising is *abstract direct representation*, which will be discussed in section 2.1.4. Darwin's theory of evolution is a good example of abstract direct representation, since Darwin did not (primarily) study models in order to arrive at his theory.

The role that is ascribed to scientific models differs from field to field, and changes over time. Only decades ago, logical positivists and logical empiricists were convinced that a scientific theory is to be construed as a hypothetico-deductive system. According to this *syntactic view*, theories are construed a set of sentences in a first order logical system. In this view, models are merely conceived of as helpful for understanding what a theory says, as convenient implementations of the theories. (Frigg and Hartmann 2017) The syntactic account was heavily criticised, and had to make place for the *semantic view* of theories. According to the proponents of this view, theories are non-linguistic entities that may be identified with a certain class of models. In this view, models no longer play an auxiliary role, models now "occupy centre stage." (Van Fraassen 1980, 44) There are many diverging versions of the semantic view. An example of a strong version of semanticism is the one advocated by Bas van Fraassen, who views upon a theory as a family of models with certain parts specified as candidates for direct representation of observable phenomena (Van Fraassen 1980, 64). A much weaker version of the semantic view is given, for instance, by Stephen Downes (1992), who proposes that the strongest claim he can make is that models are important in scientific theorising.[4]

[3]Example and wording suggested by Hughes. (Hughes 1997)

[4]One can argue about whether or not this is still a semantic view of theories, but Downes himself calls his standpoint a 'deflationary semantic view'. (Downes 1992)

Another major divide in the views upon models concerns the representational function of the model. A model can be taken to represent a certain phenomenon in the real world, but it can also be seen as a realisation of a certain mathematical theory. The latter view originates from logic, following Tarski's characterisation: "A possible realisation in which all valid sentences of a theory T are satisfied is called a *model* of T." (Tarski 1953, 11, cited by Suppes 1960, 289, emphasis added) These two accounts of what a model is are not mutually exclusive. In fact, Suppes (1960) argues that Tarski's concept of a model can be taken as fundamental also to models in the empyrical sciences. He contends that, although in the empirical sciences the use of models differs from their use in logic, the meaning of models is the same in both. Suppes' account received much critique and started a debate about what models really are. We turn to this question in the next section.

4.1.2 Ontological status of models

As mentioned earlier, there are many different kinds of models. As to the question of what ontological status models have, many different suggestions have been made by different scholars. Because the word 'model' has been used in so many ways, and varies from practice to practice, Peter Godfrey-Smith (2009, 4) argues that it is not a good idea to discuss the question of what models are. However, even though we might agree that this should not be the core of the discussion around models, it will be interesting to see the variety in opinions. This section gives an overview of the various conceptions that are alluded to in the philosophy of science regarding the ontological status of models.

Mathematical models

Models as mathematical structures is a prevailing account among proponents of the semantic view of theories. As indicated earlier, Suppes sees models as set-theoretic structures. Slightly differing from this is the view of models as state-space structures, a position held amongst others by van Fraassen and Lloyd. (Van Fraassen 1980, Lloyd 1994) This mathematical conception of models has been heavily criticised, mainly on the ground that this view fails to account for many types of models used in science that are not mathematical structures. Downes (1992), for instance, attracts attention to physical models, like the biological model of a cell. These kind of models cannot be grasped in terms of a mathematical structure.[5] Furthermore, this so-called structuralist view of models is said to diverge from the actual practice of modelling: the view fails to account for how models are actually constructed, and for how they are investigated. (Morrison 1999)

Physical models

Next to theoretical models, physical objects can count as models, regularly referred to as 'material models'. Examples of these type of models are scale models used in architecture, the metal model of DNA fabricated by Watson and Crick, and model organisms used in the life sciences. (Frigg and Hartmann 2017)

Models as fictions

Oftentimes it has been suggested that models are some kind of fictional entities. Many examples of models in science indeed are not physical objects, but rather seem to exist in the scientist's

[5] "A set of sentences may describe the cell model, but the cell model does not satisfy the description in any specifiable sense. The notion of satisfaction is a technical term from meta-mathematics with no correlate in many cases of scientific model construction." (Downes 1992, 146)

mind. The frictionless planes of physics and infinite populations of biology simply cannot exist in the real world, they are made up in order to better our understanding of phenomena in the real world. Godfrey-Smith and others have suggested that models are fictions that would be concrete if they were real. (Godfrey-Smith 2009, Suarez 1999, Frigg 2010) Another interesting suggestion is that models can be compared to fictions in literature and have a similar status.[6]

Alternative descriptions

Although the previous three kinds of models prevail in the literature, there are yet alternative views regarding the ontological status of models. Mary Morgan, for instance, defines models of economics as mixed instruments consisting of both logical –or mathematical– structure and a narrative. This narrative, or story, has a double role of explaining the model and simplifying reality. In her view, models describe and aid in understanding of the real world by telling stories. (Morgan 2001)

Another appealing approach is that of Paul Teller, who remarks that anything can be used as a model. Rather than narrowing down the possibilities of what counts as a model by looking at its ontological status, we should be guided by use. If some thing is used as a model for something (else), it counts as a model. There is no further essence to a model, there are no intrinsic features. (Teller 2001) For example, a spiral staircase can be a model for a strand of DNA, and in another intended use it may count as a model for some spring.[7]

Now that we have an idea of what models are, can be, or are taken to be[8], we may have a closer look at how models relate to the real world.

4.1.3 Model-world relationship

One of the many debates concerning the practice of modelling regards the relationship between the model and the real-world phenomena that it is taken to represent. This leads us to the interesting but hard question: *in virtue of what* is a model a representation of the real-world phenomena? In this section we explore the suggestions made in the literature of philosophy of science.

Description

Abstract, mathematical models can be described by a set of sentences, the 'model description'. In the eyes of some philosophers, for instance Suppes' (1960), scientific models are the same as logician's models, and can therefore be said to *satisfy* their model description. Ronald Giere criticises this view as being too narrow and not accounting for all scientific models. He gives a more general picture of the relation between the model description, the model and the world, represented in Figure 4.1. (Giere 1988, 1999) According to Giere, the model description, which can take the form of statements, equations or diagrams, *defines* the model. Weisberg, a decade later, criticises in turn Giere's account as too strict, because oftentimes model descriptions are not so specific that they can single out a unique model. He therefore suggests that the relation between model description and model is one of *specification*, a weaker relationship than satisfaction or definition that still accounts for 'picking out' the model. (Weisberg 2007a, 12)

[6]For a more elaborate discussion of this we refer to Roman Frigg's work. (Frigg 2010)
[7]Example taken from Suarez 1999.
[8]A working definition of 'model' will be given in section 4.1.5.

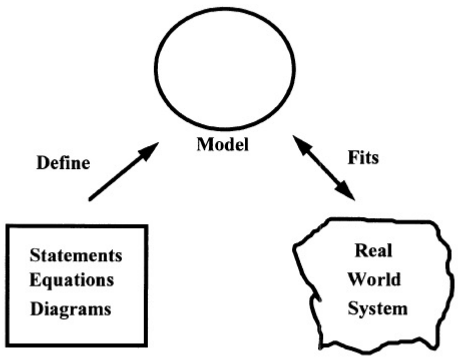

Figure 4.1. (Giere 1999, 55, figure 9)

Giere's picture of the relations between model description, model and the real-world system provides a good starting point for our discussion of the relationship between the latter two. Models somehow fit or represent some real-world phenomenon, but what precisely is the relationship between the two? In virtue of what does the model represent the phenomenon?

Similarity

Giere's answer to the previous questions is *similarity*. Other candidates have been isomorphism (Van Fraassen 1980) or partial isomorphism (Da Costa and French 2003), but these have been criticised as not being general enough to account for all scientific models.[9] Moreover, in practice it turns out that even in mathematised sciences often (partial) isomorphism is too strong a requirement. (Lloyd 1994)

Far the most promising and most discussed relationship then, is some form of similarity. However, assessing the relationship in terms of similarity has led to much criticism in the philosophy of science. Goodman (1970) calls it too vague; after all to a certain extent everything can be said to be similar to everything. A response by Giere is that the model is similar to the target system in *relevant respects* in a *high enough degree*. (Giere 1988, 81) Teller adds to this that both the aspects in which the model and the phenomenon agree, and those in which they differ, play a role in the assessment of the similarity. However, this is still relatively vague: we do not have a general account of what counts as *relevant similarity*. Teller remarks that it is senseless to look for such a general account, because what counts as relevant similarity will always depend on the details at hand. The aim, the by the researcher intended applications and the context will determine what counts as relevant.[10] (Teller 2001, 401-2) Use of the models is a key issue. This pragmatic-contextual dimension of similarity is best explained via the often discussed map analogy.

Map analogy

An ordinary city map has a lot in common with a model, and can therefore be used as an instructive analogy along whose lines the debate about models may progress. Like models, a map

[9]Another critique, voiced by Mauricio Suarez, is that isomorphism is only structural identity. Structural identity, a limiting case of structural resemblance, will never be sufficient nor necessary for representation. Peter Achinstein gives a similar argument. See for details of these arguments either Suarez 1999, 77-9, or Achinstein 1964, 28,38,41. Furthermore, there are problems regarding to what the model has to be isomorphic to. A real-world phenomenon as it is, is not always ready to be compared. Oftentimes the phenomenon has to be prepared, data has to be extracted and the raw data has to be made fit for comparison. By means of 'data reduction' and 'curve fitting' one can create so-called 'data models' (Suppes 1960) that are fit for comparison; other related notions are 'prepared descriptions' by Cartwright, 'appearances' or 'surface models' by Van Fraassen, and 'parametrised target systems' by Weisberg. Once such a data model has been created, it can be compared to the (predictions of) the constructed model. A more thorough discussion of this issue is outside of the scope of this thesis.

[10]Recall that this was also repeatedly argued in chapters 2 and 3, in particular in sections 2.4, 2.6, 3.2.2 and 3.3.

is a partial representations of something in the real world. An object that represents another object. Maps demonstrate only limited accuracy and represent in virtue of only one or more specific aspects of similarity, in this case spatial similarity between the map and the region it maps. We want to know how the streets are located relative to each other, but we usually do not need to know their elevation, or superfluous details like colours of buildings. Maps, like models, are socially constructed in accordance with criteria, interests and conventions among their constructors and users. They are context-sensitive, depend on the specific use they are intended for and the era in which they are produced. The similarities between the map and mapped region are relevant to the user: for their interests spatial relations are most important.

We can compare this example with the map of, for instance, a metro system of a big city. For users of a metro map, exact locations of metro stops and distances between them are not of utmost importance. For the intended use of these kind of maps, travelers get by with just bold approximations of the locations. What matters most now, is that the connections between metro stops are represented correctly: the user wants to know the order of the stations, and where to change lanes. Yet another example is the diagram of some electrical circuit. Here spatial relations do not matter at all for getting the right information out of the diagram; representing the connections between parts of the circuit is the sole purpose.[11] What these examples show is that the intended use, and therefore the user –the interpreting subject– is of vital importance.

Van Fraassen stresses the importance of the interpreting subject in the model world relationship even more. He argues that representation is always *representation-as* (in respect to some representational system) and is a function of the context of use. The dyadic model-world relationship discussed sofar has therefore to be seen as a triadic relationship between model, world and user.

Properties

Having discussed the nature of the representational relationship between model and the real-world phenomenon, we may now ask how a possibly abstract object may be similar to some real-world phenomenon. This is a question that has been raised by critics of the similarity approach. Commonly this question turns into the question of a comparison of properties, wherein one wonders if it is possible for abstract objects to have properties. Hughes, for instance, argues that it cannot be the case that an abstract object resembles a concrete object, as they cannot possess similar properties simply because of their different nature. (Hughes 1997, 329-30)

Other philosophers have suggested ways out of this. Following Giere, we may say that properties are *ascribed to* the abstract systems. (Giere 1988, 78) Another interesting approach is that of Roman Frigg, who compares models to literary fictions, and talks about properties that we are *entitled to imagine*. (Frigg 2010) To capture all possible versions of connecting properties to an abstract concept, we may say that the properties are *associated with* the model. (Godfrey-Smith 2009, 6)

4.1.4 Learning from a model

Now that we have seen that many scholars suggest that use, and, more specifically, *intentional use*, plays an important role in models with respect to their representational ability, we may have

[11] In these examples we are following the discussion of the map analogy by Giere and Teller. (Giere 1999, Teller 2001) See for a criticism of this analogy Hughes 1997.

a look at the actual practice of modelling. This will serve as a starting point for the discussion of how we can learn from models.

Practice of modelling

There are two[12] great accounts of the practice of modelling, provided by Hughes (1997) and Weisberg (2007a) respectively. They are to a large extent similar, and compatible, and can therefore be presented as one. We will use Hughes' DDI account, standing for Denotation, Demonstration and Interpretation, as a basis upon which to discuss insights presented by both authors.

The key insight with respect to the practice of modelling, is that the problem is represented differently, if possible simplified, and thereby we change the question and solve the problem within the new framework. Afterwards, the solution has to be translated back to the original situation. This is the core of the DDI account of denotation, demonstration and interpretation.

Denotation. The first stage of modelling is that of denoting or representing some target system[13]. The model is supposed to somehow *stand for* the targeted phenomenon, the relation between which was discussed in the previous section.

The current phase is one of the model construction. In this stage the relative intentions of the modellers play a major role, which are referred to by Weisberg as 'the construal of the model'. (Weisberg 2007a, 14-6) The construal of the model consists of several aspects:

- The *assignment*, consisting of a specification of what is to be studied, and a mapping between model parts and parts of the real-world phenomenon.
- The *scope*, then, determines which aspects of the phenomenon are to be represented by the model.
- The *dynamic fidelity criteria* describe the error tolerance, i.e., how close the output of the model has to be to the empirical data.
- Finally, the *representational fidelity criteria* match the internal structure of the model with the internal (causal) structure of the target system. This ensures that the output of the model is close to the data for the right reasons.

In this first stage lies the essential difference between modelling and abstract direct representation (mentioned in the introduction of this section). When a researcher chooses the strategy of modelling, she does not necessarily start by looking for patterns in data, instead she chooses to construct a model and investigate its properties to arrive at an answer. As Weisberg defines, modelling is "the indirect theoretical investigation of a real-world phenomenon using the model." (Weisberg 2007a, 4) This stands in contrast to abstract direct representation, in which one does not take refuge to an auxiliary model for the investigation. Weisberg gives the example of Mendeleev's construction of the periodic table. The first stage was abstraction of the raw data, resulting in a representation of the properties and relationships that are deemed relevant. The next and last stage is the analysis of this representation. As can be expected, in abstract direct representation we do not need the extra step of interpreting the results, and therefore its process consists of one stage less than the process of modelling.

[12] At least.

[13] Recall from section 2.2 that by 'target system' we mean to refer to some situation, object or phenomenon in the (physical) reality. For instance, the target system may be an atom, a ball rolling down a plane, a gas, or knowledge or belief. The target system will be represented by a formal system, a model system that has been idealised and abstracted in certain ways.

Modelling, however, consists of three stages. Once a model is constructed, it is time for the second stage of modelling: investigating the model's properties.

Demonstration. In the second stage the model is analysed, and if necessary refined by further specifying its properties. As Hughes notes, a model is not just a representation of some phenomenon, it also consists of some kind of 'internal dynamics' whose effects we can investigate. These dynamics make it possible that novel facts come about when examining the theoretical representation. Hughes gives the example of Aragon's spot (the bright spot at the centre of the shadow of a small disc), the existence of which was entailed by the wave theory of light by Fresnel. (Hughes 1997, 332)

Hughes argues that mathematics, with its deductive resources, has an internal dynamic that is very useful in physical sciences, and is often used because of this. However, next to abstract models, physical models also have internal dynamics that can be useful in aiding our understanding. Computer simulations are a relatively new, frequently used medium that bring their whole own internal dynamics. Hughes contends that all such different modes of representation contain resources that enable us to demonstrate new results. (Hughes 1997, 332)

Interpretation. The third and last stage of modelling can be seen as consisting of two steps: the assessment of the relationship between model and world, and, if appropriate, the translation of the properties and solutions found in the model to the target system.

As Weisberg mentions, the third stage of interpretation does not always take place. The model might not be sufficiently similar enough to the real-world phenomenon to give interesting input to the targeted problem, but also the model itself may be the sole object of study. An example of this are the biological models of three-sexed sexual reproduction of organisms. Of course these models do not have counterparts in the real world, but looking further than the actual possibilities can be really instructive for explaining two-sexed sexual reproduction. As Eddington wrote: "...the contemplation in natural science of a wider domain than the actual leads to a far better understanding of the actual." (Eddington 1929, 266, cited by Weisberg 2007a, 19) In these kind of cases modelling might only consist of the first two stages, and will not involve an assessment of the model-world relationship.

If the model-world relation is assessed to be correct with respect to the fidelity criteria mentioned in the first stage of modelling, that is, if the model is similar enough to the real-world phenomenon for present purposes, the (partial) solutions and suggestions by the model can be translated in terms of the real-world phenomenon. This is the final step of interpretation.

How to learn from models

It is claimed that impossible scenarios can be fruitfully used and often occur in science (Godfrey-Smith 2009, 13), but the question remains how we can extract knowledge from such abstract, possibly even unrealistic models. The previous account of denotation, demonstration and interpretation has given an idea of how we can learn from modelling in practice. Mary Morgan and Margaret Morrison (1999) have argued that we chiefly learn from modelling in the first two stages, the stages of construction and manipulation. They claim that, because models are partially independent both of theory and of the real world, they can function as autonomous agents and can therefore serve as instruments of exploration in both theory and reality.[14] (Morgan and Morrison 1999, 10,24-5) Especially the internal dynamics of models discussed by Hughes play a crucial role

[14]This is, however, not their only function, as they also serve to represent the target phenomenon. (Morgan and Morrison 1999, 24)

for our prospect of learning from them.

Godfrey-Smith remarks that what we learn from a model always takes the form of a conditional: if the problem were to be like this, the solution would be like that. Furthermore, he incites us to distinguish between the insider and outsider perspective. When in a particular field of science fictionalising with the use models becomes more systematic, he argues, a tradition comes into existence in which the fictional entities are studied as topics in their own right. Because the scientists of this particular field get used to discussing these fictional models, they begin to treat them as real objects. Godfrey-Smith suggests, somewhat negatively, that then "modelers [might] achieve something useful despite themselves," and that outsiders might reinterpret their claims in terms of conditionals. (Godfrey-Smith 2009, 18-9)

Aims and functions of models

"Models can instruct us about the nature of reality only if we assume that (at least some of) the model's aspects have counterparts in the world." (Frigg and Hartmann 2017)

A central aim when constructing a model is faithfully representing at least some of the parts of the target phenomenon. If the model is related to the real-world phenomenon in the –for current intentions, in the current context– right way, the model can serve several functions. A non-exhaustive list of functions that models may have: simplifying and thereby clarifying, giving structure to the problem, detecting difference-making factors, aiding scientific imagination, exploring beyond possible scenarios, telling stories, and giving a general picture.

Models may have different functions in relation to theories. They can be superfluous additions that have some kind of hermeneutical value (as held by proponents of the syntactic fuel theories), they may be the central unit of a theory (as held by proponents of the semantic view of theories), they may be relatively independent from both theory and data and can therefore act as autonomous agents (as suggested by Morgan and Morrison), they may complement theories that do not provide an account of concrete situations (as argued by Cartwright, an example of this would be quantum field theory), they may depict theories that are too complex to handle, they may serve as a starting point for theories, and they may serve as a substitute when theories are not available (as is common in economics and biology, discussed by Groenewold 1961). Again this is not an exhaustive list of functions, but it gives a good overview of the possibilities.[15]

Relationship between models and representational imperfections

Cartwright reminds us of the inherent incompleteness of a (theoretical) model. (Cartwright 1989, 192) It is bound to leave certain aspects of the target system out, or distort some aspects, for it is impossible to grasp the full physical reality in a theoretical model. Robert Cummins contends that abstraction is inherent to the construction of some object as similar to some other object, and therefore abstraction is inherent to modelling. (Cummins 1989, 33) Also Giere remarks that models never really fit reality. Take for instance an equation describing some physical process. If we want to make this equation fit our data, we will always have to include some 'margins of error', plus minus some delta that may be specified or not. These deltas, Giere argues, can be seen as an assessment of the degree to which the equation is similar to the actual situation. (Giere 1999, 50) Also Teller (2001) argues that the perfect model does not exist: inevitably some aspects of the target phenomenon will be misrepresented by the model. Furthermore, Giere notes that –as most probably comes as no surprise to scientists involved in the practice of modelling– models come in various degrees of abstraction. (Giere 1988, 79) We may conclude that the practice of

[15] Examples taken from Frigg and Hartmann 2012.

modelling is intimately connected to abstraction and idealisation. We will turn to a discussion of abstraction and idealisation in the next section.

Further connections

As can be seen from the previous paragraphs, modelling is intimately connected to both computer simulations and thought experiments. Whether or not these two are taken to be forms of models depends on the particular philosophical conviction of the scientist or philosopher of science. We will not delve into this any deeper, but contend in saying that an intimate connection between these forms of theorising exists.

4.1.5 Choices and definitions

We have seen that the word 'model' can refer to many varying concepts. We might agree with Teller that anything can be used as a model, and that whether or not it counts as a model, and what it represents, is determined by its intended use. In this thesis, however, we do not need such a broad account of a model. To avoid confusion, we will therefore define a model as an abstract (mathematical) construct[16], which will determine the use of this concept throughout the rest of this writing. It serves the purposes of this thesis best to define 'model' as 'a theoretical construct that serves to represent some target system'.

As discussed, Cartwright reminds us of the inherent incompleteness of such a theoretical model, which is bound to omit or distort aspects of the real-world phenomenon. This omission or distortion of certain aspects or features of the target system usually goes by the name of 'idealisation' or 'abstraction'. These words may be used both to describe the process of idealising or abstracting the model, or to the resulting ideal or abstract model. This usually happens without confusion, and wherever ambiguity lurks, we may resort to using the verbs 'idealising' and 'abstracting', and the nouns 'ideal model' and 'abstract model'.

Having done the preparatory work of investigating modelling in science and positioning ourselves with respect to this information by defining our vital concepts precisely, we may now discuss some work on abstraction and idealisation in the philosophy of science.

4.2 A variety of definitions concerning idealisation

As discussed in the introduction of this chapter, a selection has been made of authors that present varying and interesting views with respect to the topic of idealisation in science. These philosophers of science have written about idealisation in the past four decades, and will be presented chronologically because they often discuss ideas that have been raised and published prior to their publication. The philosophers of science tol be discussed in this section are Ernan McMullin, Nancy Cartwright, Martin Jones, Michael Weisberg, and Martin Stokhof and Michiel van Lambalgen.

4.2.1 Ernan McMullin

In his essay *'Galilean Idealisations'* of 1985, Ernan McMullin discusses several idealisation techniques that Galileo employed in his time, and places them under the heading 'Galilean idealisations'. An idealisation, for McMullin, is the deliberate simplification of something complex in order to improve our understanding of it. Galilean idealisations are not necessarily invented by

[16]We should note that this does not make the model itself a mathematical entity, it merely means that it can be mathematically described. (As remarked in Godfrey-Smith 2009, 5)

Galileo nor did he necessarily play an important role in their development, but they have been christened 'Galilean' because these forms of idealisation played an important role in shaping "the new science." (McMullin 1985)

McMullin distinguishes between two ways in which Galilean idealisation can proceed: construct and causal idealisation. *Construct idealisation* involves a conscious simplification of the conceptual representation of the phenomenon in order to escape part of the complexity of the real world. In *causal idealisation* the physical reality itself is simplified. Among construct idealisations we can further distinguish between formal idealisation and material idealisation. In *formal idealisation*, features that are deemed to be relevant are deliberately simplified or omitted[17], while still retaining the essence of the target system. The simplified features ultimately have to be allowed for in a process of de-idealisation. The steps in this process of making the idealised model more specific should be suggested by the original misrepresentations and omissions. McMullin points out that if modification of the idealised model could not be derived theoretically, but only in an empirical way, then the model itself would be suspect. Such corrections would be regarded as ad hoc, and scientists will have to look for a more coherent model.

Mathematical idealisation, according to McMullin, is a type of idealisation that is necessarily formal. However, because of its special place in (philosophy of) science, it is discussed separately. Mathematical idealisation is the process of laying down a mathematical formalism on the physical reality. This type of idealisation was used long before Galileo's time, but is connected to his name because of the impact that his geometrisation of the science of motion had.[18] Mathematical idealisation is also discussed and criticised by Pierre Duhem, although he gives it the name 'abstraction'. (Duhem 1962) In order to distinguish her own sense of abstraction from this type of abstraction, Cartwright refers to it as the kind of abstraction that leads to 'symbolic representation'. McMullin asserts that, while the 'Book of Nature' is possibly not written in the language of mathematics[19], mathematics provides the syntax or grammar. Physics is not reduced to mathematics, but mathematics can be expanded to serve as the syntax of physics in any case imaginable.[20] According to McMullin, mathematical idealisation is always formal idealisation, but not the other way around. The stimulus/response model of the psychologists is an example of a non-mathematical formal idealisation. (McMullin 1985, 258)

Material idealisation, the other type of construct idealisation, is the matter of leaving unspecified those features deemed irrelevant to the current question. Whereas in formal idealisation features of the target system are misrepresented, in material idealisation further specification of certain for the inquiry at hand irrelevant features are being laid aside for the moment. Both kinds of construct idealisation are employed for the purpose of a better understanding, and according to McMullin the process of de-idealisation serves as a warrant of the model. A difference however is that this process is fairly straightforward in the case of formal idealisation, but in material idealisation the initial omissions rather serve as a source of suggestion. (McMullin 1985, 254-264) McMullin purports that construct idealisation of both forms is truth-bearing "in a very strong sense." (McMullin 1985, 264) He argues that the idealised model has to fit the target system in the physical reality well in order to yield good results and predictions. The process of self-

[17]According to Martin Jones, McMullin here intends to refer to 'representation as absent' instead of 'not at all mentioning'. (Jones 2005, 189)

[18]It is mathematical idealisation that Husserl has in mind when he speaks of 'Galilean idealisation', according to McMullin. (McMullin 1985, 254)

[19]Galileo suggested that God had written two books: the Bible and the Book of Nature. The former could be interpreted by theologians, the latter by mathematicians as it was written in the language of mathematics. Whether or not the Book of Nature indeed is written in the language of mathematics has been debated vigorously.

[20]"It would be hazardous today to argue that there are causal factors at work in the natural world that are inherently incapable of being grasped in mathematico-physical terms." (McMullin 1985, 254)

correction or de-idealisation he views upon as "the best testimony of truth." (McMullin 1985, 264) The observation that construct idealisation works well for the physical sciences, that it produces verifiable predictions, is to McMullin a strong argument for scientific realism.[21,22]

Under the heading 'causal idealisation', McMullin distinguishes experimental from subjunctive idealisation. *Experimental idealisation* is the matter of isolating causal lines in a real experimental setup in order to study its influence on the target system. The features that thereby are deliberately lessened are treated as impediments that (for now) hinder understanding of the natural tendency of the target system. This results in an understanding in the form of ceteris paribus laws, stating what would happen if all other factors are held constant. McMullin remarks that this technique of isolating causal lines does not work so well for biology and psychology, in which the target system is a complex organism or a social group. (McMullin 1985, 268) McMullin calls experimental idealisation most typically Galilean, for it was Galileo who had the insight that causally complex systems can be understood[23] by first isolating causal lines and thereafter combining them.

The other form of causal idealisation is *subjunctive idealisation.* This is the conceptual type of idealisation in which one focuses on one causal aspect by means of a thought experiment. It asks what would happen under 'normal' circumstances, or under circumstances in which all other causal factors are held the same. This question is answered by taking refuge to intuition (scientific intuition or intuition from everyday experience). Its goal may be exploration of the role of the isolated causal factors, what the additional relevant causal factors are, and how these are related to each other. Like mathematical and experimental idealisation, this technique was far from new in Galileo's time. However, it played a central role in Galileo's mechanics. (McMullin 1985)

4.2.2 Nancy Cartwright

In her book, *'Nature's capacities and their measurements'* of 1989, Nancy Cartwright distinguishes between idealisations and abstractions. When philosophers speak of 'idealisation', she notices, they usually refer to a blend of both. Cartwright suggests to reserve the term '*idealisation*' for the process of misrepresenting some relevant but inconvenient features of the phenomenon under study. These factors cannot be just deleted, for a model must say something about all relevant factors. As an example, Cartwright takes a ball rolling down a plane. To calculate the motion of the ball, you must have some value for the forces in the three orthogonal directions; if you were to simply omit one of them, you could not study the motion. The specific values of these forces may be misrepresented or fixed, but cannot be left out. In contrast, *(Aristotelean) abstraction* in Cartwright's terms refers to the process of genuinely taking away or 'subtracting' of all that is irrelevant to the inquiry at hand. These irrelevant aspects may be certain features of the phenomenon, circumstances and material. An example of this would be the colour of the ball of which we study the motion.

Idealisation and abstraction for Cartwright are both processes of thought, of changing relevant features and subtracting irrelevant features, respectively. In the former, the notion of 'departure from the truth' makes sense as all the misrepresented factors have in reality some specific value, hence a departure from truth could be calculated. In abstraction however, speaking of a smaller

[21] A moderate version of scientific realism, as developed in McMullin 1984.

[22] In the essay *'Galilean Idealisations'* McMullin counters Cartwright, who for a major part denies that idealisations in science are truth-producing. McMullin acknowledges that there are laws in physics (for example, in quantum mechanics) that lack the possibility for theoretical self-correction, but denies that this holds for all laws of physics. In *'How the Laws of Physics Lie'*, only laws derived experimentally escape the fate of being named a lie. (Cartwright 1983)

[23] "Only understood," writes McMullin. (McMullin 1985, 265)

or larger departure from the truth makes no sense, as features are genuinely subtracted and therefore a hypothetical calculation of the departure from truth is not at all possible.

However, according to Cartwright, idealisation and abstraction are intimately connected[24]: the aim of idealisation is constructing an ideal model upon which we can establish an abstract law. It should then come as no surprise that when speaking of idealisation Cartwright mostly refers to models, whereas in connection to abstractions she mainly refers to laws. With the term abstraction, Cartwright has in mind primarily the dropping of 'ceteris paribus' in ceteris paribus laws. That is, where a (ceteris paribus) law arrived at via an ideal model would say something like 'X is the case under the ideal circumstances that I', the abstract law would only say 'X is the case'. In other words, an abstract law wins in generality compared to the original law.

Cartwright's notion of *'Aristotelean' abstraction* is to be distinguished from Duhem's notion of abstraction[25]. She remarks that abstract properties arrived at by Aristotelean abstraction are realistic (but possibly counterfactual) properties, while those mathematical symbols suspected by Duhem do not correspond to anything in nature. (Cartwright 1989, 197) Because of the focus on laws in abstraction, and the sole focus on models in idealisation, Cartwright's terminology is sometimes hard to compare to that of other philosophers. The focus in both concepts, however, lies on the isolation of causal factors. Cartwright calls causal idealisation the "most relevant" aspect of Galilean idealisation, and in her work she explicitly focuses on causal abstraction. (Cartwright 1989, 185, 198)

4.2.3 Martin Jones

In response to the variety of usages of the words 'idealisation' and 'abstraction'[26], in his essay *'Idealisation and Abstraction: a Framework'*, Martin Jones (2005) purports to provide a framework upon which discussion of such representational imperfections can take place. With this aim he gives one necessary condition for idealisation and one for abstraction. Whereas *idealisation* is supposed to necessarily involve misrepresentation, *abstraction* is taken to require mere omission. This makes idealisation and abstraction mutually exclusive: something counts as abstraction only if it involves omission without misrepresentation. Jones makes a case that idealisation needs some approximation of truth, but speaks against making this a necessary condition. He argues that idealisations typically make a model simpler or more tractable, and that one typically aims at relevant features when idealising. However both of these considerations do not give us any further necessary or sufficient conditions for idealisation. Abstractions, on the other hand, can according to Jones (in line with Cartwright's argument) not approximate the truth in any sense, since they keep completely silent about certain features of the system. Abstractions inherently contribute to simplicity, and although they usually involve omission of irrelevant features, Jones (2005) also makes a case for mere omission of relevant features.

Because of failing attempts to extract more necessary or sufficient conditions, Jones (2005) is very critical of the approach of defining idealisation and abstraction in terms of a collection of necessary and sufficient conditions. Importantly, Jones defines the terms abstraction and idealisation with respect to the truth, not with respect to our conception of the truth. It is therefore possible to idealise or abstract without being aware of it, or to discover that something has been an idealisation.

[24]Even, "idealisations would be useless if abstraction were not already possible." (Cartwright 1989, 188)

[25]As discussed in the previous paragraph under 'mathematical idealisation'.

[26]Jones both criticises the conflation and confusion around these concepts, the lack of homogeneity in their uses, as well as does an appeal to common usage. This seems to be contradictory.

4.2.4 Michael Weisberg

Michael Weisberg argues in his article *'Three Kinds of Idealisation'* that idealisation in science can be distinguished to be of three types. These three kinds of idealisation reflect different practices in science, are connected to different scientific goals or representational ideals, and are motivated differently. (Weisberg 2007b)

The first kind of idealisation that Weisberg distinguishes is *Galilean idealisation*, for a discussion of which he refers to McMullin's analysis[27]. In Galilean idealisation, distortions are consciously introduced in our theories and models in order to, according to Weisberg, make them computationally tractable. Therefore Galilean idealisation is mainly important in the investigation of computationally complex systems. The ultimate goal of Galilean idealisation is to allow for these distortions later on in the process of de-idealisation, in order to give a representation of the target system as complete as possible. Although Weisberg does not deem a complete representation of real-world phenomena possible, completeness provides an aiming point and simultaneously sets up a scale upon which to judge the idealised model. The temporary distortions introduced in Galilean idealisation are justified pragmatically. (Weisberg 2007b, 640-2, 655)

The second kind of idealisation is called *'minimalist idealisation'*. This is the practice of constructing theoretical models that include only the most important causal factors present in the target system. Because only the core causal factors are included (those factors that make a difference to the phenomenon under study), this type of idealisation lends oneself well to facilitate scientific explanation. Minimalist models can for example be used for pedagogical purposes, to test general ideas and to aid understanding. Weisberg refers to Cartwright's account of abstraction as a form of minimalist idealisation. (Weisberg 2007b, 642-5, 655)

The third and last type of idealisation is *multiple-models idealisation*, by Weisberg defined as follows: "multiple-models idealization (hereafter, MMI) is the practice of building multiple related but incompatible models, each of which makes distinct claims about the nature and causal structure giving rise to a phenomenon." (Weisberg 2007b, 645) MMI is the answer to the problem faced by scientists who in their research are governed by multiple goals. For example, scientists may desire to achieve precision, simplicity, generality and accurate predictions simultaneously. Due to the complexity of the world and our cognitive and computational limitations, these representational ideals may not be jointly accomplished in one single model. To achieve all of our scientific desiderata at the same time, we may need to construct multiple models. An example of a field of science in which MMI is a common practice is metereology. As Weisberg notes, the United States National Whether Surface uses three complex models of global circulation patterns in order to model the weather faithfully and to obtain accurate predictions. As can be expected, MMI does not have one single scientific goal, nor is there a single motivation for this type of idealisation. The practice may or may not be permanent, it may or may not be justified pragmatically. (Weisberg 2007b, 645-8,655-6)

According to Weisberg, the three kinds of idealisation are not competitors, but distinguished practices that reflect different strands of scientific inquiry. Different goals ask for different models. For example, a black box model that has great predictive power for unknown reasons, is justified easily when the goal is to predict with great precision and accuracy, but when the aim is scientific explanation it will leave us empty-handed. Models and their idealisations, therefore, are to be judged with respect to their scientific goals or representational ideals.

[27] McMullin 1985, as discussed in section 4.2.1.

4.2.5 Martin Stokhof and Michiel van Lambalgen

In their article *'Abstractions and idealisations: the construction of modern linguistics'* Martin Stokhof and Michiel van Lambalgen propose yet another definition of idealisation and abstraction. To them, *'abstraction'* refers to the kind of distortion that often occurs in models in the physical sciences, sciences of quantitative nature, whereas the term *'idealisation'* is reserved for misrepresentation of a more qualitative nature. The dichotomy is based on three main differences that seem to be interconnected: parameters being fixed versus features being ignored, quantitative versus qualitative features, and the motivation being primarily methodological and practical versus primarily ideological and theoretical. The main aim of the article is to give a critique on the practise of formal semantics by revealing the distinction between physical sciences and the formal semantics: misrepresentations in models of the former are what Stokhof and Van Lambalgen call abstractions, whereas misrepresentations in models of the latter are idealisations. (Stokhof and Van Lambalgen 2011)

An important consequence of the distinction is, that because of the quantitative nature in abstraction, some 'bridge theory' is not needed: the parameter misrepresented is still present and can thus in principle easily be incorporated, whereas idealisation calls into being the further epistemological question of how the idealised model is to be related to the target phenomenon. Stokhof and van Lambalgen argue that in formal semantics this bridge theory is not given, and there is little chance that we will be able to find one. This is also related to the connection to experiment: whereas in the physical sciences experimental results give an impartial judge, models in formal semantics cannot be tested in such way. Consequently their success depends on the way we define it to be, which is (at least partially) defined by our choice of methodology. (Stokhof and Van Lambalgen 2011) Stokhof proposes that formal languages can serve as an instrumental model of natural language, because they can be useful tools for the study of language as occurring in the real world.[28] (Stokhof 2012)

Differences in terminology

As can be seen from the previous discussions of terminology used by various philosophers of science, their conceptions vary with respect to multiple dimensions. This is graphically displayed in the table on the next page. An elaborate discussion of this follows in the next section.

[28] More precisely, Stokhof proposes that, *instead* of seeing formal languages as a model of some natural language, we see them as useful tools for exploration of the concept of natural language. This is due to a different use of the word 'model'.

Table 4.1: Uses of terminology around idealisation in science

Name author	Ernan McMullin (1985)					Nancy Cartwright (1989)		Martin Jones (2005)		Michael Weisberg (2007)			Martin Stokhof (2011/16)	
Concept	Mathematical Idealization*	Formal Idealization	Material Idealization	Experimental Idealization	Subjunctive Idealization	Idealization	(Aristotelian) Abstraction	Idealization	Abstraction	Galilean Idealization**	Minimalist Idealization	Multiple-Models Idealization	Abstraction	Idealisation
Goal	symbolic representation	completeness, verifiable predictions	scientific explanation	scientific explanation	scientific explanation, exploration	scientific explanation, abstract laws	scientific explanation, generality			completeness	scientific explanation	multiple goals, often including accurate predictions	tractability	
Justification	pragmatic	pragmatic	scientific explanation	pragmatic	scientific explanation	pragmatic	scientific explanation			pragmatic	scientific explanation	mixed	pragmatic	ideological
Misrepresentation or omission	symbolic representation	misrepresentation	omission	both, isolating causal lines	omission	misrepresentation	omission	misrepresentation	omission	both	omission	both	misrepresentation	omission
W.r.t. truth or conception								truth	truth	conception	conception	conception	conception	conception
Relevance of features	relevant	relevant	irrelevant	irrelevant	irrelevant	relevant	irrelevant	likely relevant	if interesting relevant	both	irrelevant	both	irrelevant or too complex	relevant?
Kind of features	quantitative?	both quantitative and qualitative	all kinds	causal	causal	focus on causal	focus on causal			quantitative, causal	causal	all kinds	quantitative	qualitative
Permanence		non-permanent				non-permanent	non-permanent			non-permanent	permanent	mixed permanence		
Particularities					in thought, ceteris paribus								no bridge theory needed	bridge theory needed

* Recall that this has been referred to by Duhem as abstraction, and by Husserl as Galilean idealization.
** Recall that Weisberg refers to McMullin for an analysis, and thus encompasses the five types of idealization mentioned under McMullin's name. Remark that Weisberg and McMullin's conception of Galilean Idealization is broader than Husserl's.

4.3 Discussion

The table on the previous page gives an overview of the different kinds of idealisation that have been discussed. For example, Weisberg's minimalist idealisation is characterised as the pragmatically motivated permanent omission of irrelevant causal features of the phenomenon with respect to our conception thereof with the aim of scientific explanation. The variety of concepts discussed can be distinguished among eight dimensions: goals, justification, misrepresentation versus omission, with respect to the truth or our conception thereof, relevant versus irrelevant features, kind of features, permanence, and further particularities. These distinctions will be discussed and ultimately we will argue which distinctions are most suitable to underlie our investigation of abstractions and idealisations in formal epistemology.

Goals and justification

As can be seen from the table, justifications for idealisations or abstractions in a model can differ in nature connected to different goals. Here we take 'goal' to refer to the object or desired result of the process of idealising or abstracting.[29] With respect to the nature of justifications we will distinguish between pragmatic and non-pragmatic justification. With a pragmatic justification we mean a methodological or practical consideration, with non-pragmatic justification we refer to ideological or theoretical considerations. For instance, McMullin's formal idealisation is justified pragmatically with the goal of completeness (by the process of de-idealisation) and giving verifiable predictions. Material formalization according to McMullin is justified non-pragmatically with the goal of scientific explanation. Like McMullin's formal idealisation, Cartwright's idealisation is justified pragmatically. The goal of Cartwright's idealisation, however, is to generate abstract laws, with the ultimate goal of scientific explanation. Cartwright's abstraction, then, is justified non-pragmatically with the goal of scientific explanation and generality. This goal of generality is connected to the fact that Cartwright has abstract laws in mind in which the ceteris paribus conditions are dropped. Generality may also be one of the goals of multiple-models idealisation. Other possible goals of multiple-models idealisation may be for instance yielding accurate predictions, completeness, exploration and scientific explanation. Because of these multiple goals, the justification of multiple-models idealisation is mixed.

Looking at the table we may remark that Weisberg's Galilean idealisation, which was to encompass all of McMullin's idealisation types, is justified solely pragmatically with completeness as only goal. In this respect Weisberg's Galilean idealisation seems to be most like McMullin's formal idealisation, and less like mathematical, material, experimental or subjunctive idealisation. Stokhof[30]'s abstraction resembles formal or Galilean idealisation in the respect that its justification is pragmatic and its goal tractability. However, we may note that tractability itself is probably not an ultimate goal for formal or Galilean idealisation as is completeness: tractability is more of an intermediate goal that makes it easier to achieve the ultimate goal of completeness. Stokhof's idealisation is justified ideologically or theoretically, which distinguishes the concept greatly from Stokhof's abstraction. With 'abstraction' Stokhof means to refer to the kind of idealisation that happens in the physical sciences. However, as we may see from the other concepts by philosophers of science, this kind of pragmatically motivated idealisation to get more tractability is not the only kind of idealisation in the physical sciences. Non-pragmatically motivated idealisation with the goal of scientific explanation can also be found in the physical sciences.[31]

[29] This notion is closely related to Weisberg's representational ideals, and was inspired by his approach. See for a discussion of this in relation to idealisation in science, *"Three Kinds of Idealisation."*. (Weisberg 2007b)

[30] Stokhof and Van Lamblagen. For brevity and readability, we sometimes only mention Stokhof in this section.

[31] See for instance McMullin's material idealisation, Cartwright's abstraction or Weisberg's minimalist idealisation.

Misrepresenting or omitting (ir)relevant features

Oftentimes in the definitions we see the combination of either 'misrepresenting[32] relevant features' or 'omitting irrelevant features'. Stokhof's characterisation of the concepts abstraction and idealisation are herein the great exceptions: abstraction is defined as misrepresenting irrelevant or too complex features, and idealisation seems to be defined as omitting relevant features. At least, this seems to be the critique Stokhof and Van Lambalgen make against formal semantics: that the features that are being laid aside for the moment are actually relevant, that language cannot be taken apart from actual language use and the inconvenient properties that this entails, like finity of the natural language.

Somewhat similarly, Jones argues that abstraction as mere omission is most interesting when the features that are being laid aside are relevant. However, in his discussion of idealisation, Jones comments that idealisation likely involves misrepresentation of relevant features (but this does not make for an auxiliary necessary condition to his definition). Jones also remarks that McMullin's formal idealisation probably involves no real omission but rather misrepresentation as absent. If this is true, then McMullin's formal idealisation can be seen as a kind of idealisation as characterised by Jones, and it also matches Stokhof's definition of abstraction a bit better.[33]

Defining the concept w.r.t. truth or conception thereof

Dependent on your philosophical opinion and views, you may choose to view upon a representative imperfection as deviating from either the truth of the matter, or from your conception of the target system. Choosing the former or the latter has influence on whether something counts as a representative imperfection or not. For instance, when something is misrepresented because of lack of knowledge, this may count as an idealisation for Jones, however, this does not count as an abstraction in Stokhof's sense. The difference lies in that Jones sees an idealisation as a misrepresentation with respect to the truth of the matter regarding the target system, whereas Stokhof sees abstraction as a misrepresentation with respect to our conception of the target system. In Stokhof's sense of the word, something may only count as a representative imperfection if features of the target system are deliberately misrepresented or omitted. A simple lack of knowledge of the underlying mechanisms does not make some misrepresentation an idealisation. It cannot be discovered that something was an idealisation (as is the case in Jones' framework).

Kind of features

According to Stokhof the most important distinction that underlies the difference between the concepts abstraction and idealisation is the distinction between quantitative and qualitative features that are being misrepresented or ignored. We do not really see this distinction elsewhere. McMullin expresses that formal idealisation can be both quantitative and qualitative in character, and that this depends on the respective field of science in which idealisation takes place. However, for McMullin this is not a factor upon which to distinguish different kinds of idealisation. For Stokhof it is, because the aim in his joint work with Van Lambalgen is to make the distinction between the physical sciences and formal semantics clear, entailing different standards for idealisation. For Cartwright the focus in both idealisation and abstraction lies on causal features.

[32]'Misrepresenting here refers to representing in a way that is not necessarily true to the real phenomenon, examples of this can be fixing the value of a certain parameter, or rounding up some parameter in an equation. Note that misrepresenting in this sense may in fact (incidentally) give an exactly true account (although this is highly unlikely), and it does not carry the negative connotation of 'representing in the wrong way'.

[33]But not fully: Stokhof and Van Lambalgen also count mispresentation of irrelevant features to abstraction, and confine it to quantitative features only.

Permanence

Types of idealisation can also be distinguished with respect to their permanence. Formal and Galilean idealisation are distinctively non-permanent. This stands in contrast to, for example, minimalist idealisation, which is permanent in character. According to Cartwright both idealisation and abstraction are non-permanent practices, and misrepresentations or omissions can be ultimately accounted for.

Bridge theory

One of the most new and useful distinctions that Stokhof and Van Lambalgen bring to the table is whether or not a bridge theory is needed to connect the phenomenon under inquiry to the resulting model. Their view is that idealisation encumber us with additional epistemic questions that need to be answered before we can extrapolate answers found in our formal system to our target system. Whether or not such a bridge theory is needed will be central in our approach.

4.3.1 Redefining abstraction and idealisation

From the variety of definitions that is available we will need to choose some concept of abstraction and/or idealisation, or some main lines along which to define the concepts. We will first outline some starting points.

- As we have seen from the previous sections, abstraction and idealisation depend on what we want to explain, describe or predict. It depends on the aspect of the target system that we are focusing on, it depends on our goal. Therefore, we need to make this clear every time we use this terminology.
- As we have seen from the table and the discussion of the concepts, different kinds of idealisation can be defined with respect to the truth or with respect to our conception of the truth. We will define our idealisation types with respect to the conception of the truth, as this is more convenient to work with. This way, we will be able to say if something is an idealisation or abstraction or not.
- Also relevance of the features is defined with respect to our conception: the question is whether we *deem* the features under inquiry relevant, not whether they *are*.[34] Later in the process we may find that an aspect we deemed irrelevant was relevant after all, or vice versa.
- A leading principal in our research will be whether a bridge theory is needed.

In order to use the concepts of 'abstraction' and 'idealisation' to talk about representational imperfections in epistemic logic more precisely, we need to redefine these concepts. As we have shown, there are many dimensions along which we may choose to define abstraction and idealisation. We choose to redefine them along two dimensions: a combination of the dimension of omission versus misrepresentation and the dimension of (deemed) relevant versus irrelevant features. We choose these dimensions to underlie our new definition, because it helps us stay closest to definitions used in philosophy of science, and it suits our current purposes best.

Definition 4.1. *(Abstraction, idealisation and false idealisation)*
- *A representational imperfection will be called an* ***abstraction*** *when it omits an irrelevant feature of the target system.*
- *A representational imperfection will be called an* ***idealisation*** *when it misrepresents a feature of the target system that is deemed relevant (or irrelevant, but this is usually less interesting). Note that misrepresentation here means either being represented in the model*

[34]Recall that Van Fraassen and others stressed the importance of the use and user in the evaluation of a model, section 4.1.3.

in a way that deviates from the reality of the target system, or that this deviation from reality is explicitly mentioned in the discussion that accompanies the model.

- *A representational imperfection will be called a **false idealisation** when it omits a relevant feature of the target system.*

Discussion

Instead of giving just two new definitions, we gave three. The reason for the extra definition of the 'false idealisation' is that researchers may disagree about which features of the target system are relevant or not. A researcher may present their model as an idealised model, upon which another researcher may criticise it for being a falsely idealised model: it actually omits some relevant features of the target system. As Jones remarked in section 4.2.3, omitting relevant features of the target system can be interesting for certain applications. However, in order for it to be noticed and found to be interesting, we should mention this omission in the discussion of our model.

Incorporating 'mentioning in the model discussion' in the definition of idealisation (additionally to being present in the model in a way that deviates from the reality of the target system), is also in line with the thought expressed in section 3.4.x. We should be open about the idealisations we make, so that they are up for discussion. If we do that, and we remain aware of them, they may be considered to be acceptable. This is in contrast with some of the definitions we saw in the physical sciences. Recall, for instance, McMullin's critique in section 4.2.1: if the possibilities for the idealisation are not suggested by the model itself, one might suspect the model to be not appropriate. The differences between models in the physical sciences and models in epistemic logic, as well as McMullin's remark, will be further discussed in sections 5.1.1 and 5.1.4.

If we look at the examples of representational imperfections[35] given in section 2.4, we can remark the following: Qualitative idealisations like hair colour will often be abstractions, and sometimes idealisations, in the case of qualitative features being relevant to the intended aims and application. Useful fictions like common knowledge will usually be idealisations, as they are consciously introduced. However, byproducts of our formalisation, like logical omniscience, have the danger of being false idealisations: the modeller may not mention them, as he did not introduce them consciously (and, as we saw in section 2.5, the various forms of logical omniscience is very hard to escape while using the framework of epistemic logic).

By means of this threefold definition we covered all possible representational imperfections there are, and give a useful vocabulary for speaking about them in epistemic logic. The third definition gives voice to a critique similar to that by Stokhof and Van Lambalgen about formal semantics: the critique that some relevant aspects of the target system have been left out completely, thereby changing the question so drastically that it leaves us with the extra epistemological burden of providing a bridge theory. In the same vein, we might want to say for some epistemic logical models in certain research contexts, that they cannot make any further claims than the claims about the system itself, because the model deviates too much from the reality of the target system to extrapolate results reached within it. Before the results could be used to speak about the target system, then, a bridge theory should be provided showing how the formal system relates to the target system.

We may use these new definitions to give some structure and clearly defined vocabulary in order to have a thorough discussion of the circumstances under which some representational imperfections may constitute a problem towards the practice of epistemic logic or not. This will be discussed in chapter 5.

[35] Recall that this word refers to any type of deviation from the truth of the target system.

Chapter 5

Discussion and Conclusion

Multiple epistemic logicians expressed during their interview that 'all sciences idealise'. This chapter will look at a comparison of the models used in epistemic logic and models of other sciences, and will bring together the lessons that can be learned from the previous chapters.

5.1 Learning from an ideal model

5.1.1 Models in the physical sciences compared to models in epistemic logic

"Common knowledge of rationality is not something about the real world, it's a sort of benchmark. It says: this is the way things work out when there's no irrationality in the system. If there is no irrationality in the system at all, this is how things work out. Everything else is calculated as a departure from the benchmark. It's like a no friction assumption in physical systems, or the no air resistance assumption. ... you start out with the assumption that under ideal conditions, you have no air resistance, and then from there you can go and calculate what happens with air resistance."

- Robert Aumann (EL5Q, 30)

Although also the models of other sciences idealise, and these idealisations seem to have a lot in common with the idealisations of epistemic logic, we better first investigate to what extent epistemic logical models are similar to models of the physical sciences[1], before we continue to discuss the lessons that we may learn from the literature of philosophy of science.

Similarities Both the models of the physical sciences and those of epistemic logic idealise and abstract away from many features of the target system. They do so in order to reach certain scientific goals: the goal of complete representation, scientific explanation, and so on. The practices of modelling both in the physical sciences and in epistemic logic often follow roughly the same pattern: first we change the question, then we use the internal dynamics of our model to reach some results within our system, which we thereafter interpret. Both in the physical sciences and epistemic logic this interpretation can be problematic, as the model-world relationship is not always crystal clear.

Differences One main difference between models in the physical sciences and models in epistemic logic is that the first are largely of a quantitative nature and the second of a qualitative nature. Also in the first we more easily construe experiments to check our model, this is much less the case for epistemic logic: the objects of study of the physical sciences are often more appropriate for designing experiments than the object of study in epistemic logic. In the former one usually studies real-world phenomena, in the latter the central notion is a concept which lives in human minds (knowledge). Furthermore, the agents that are studied in epistemic logic do not

[1] After all, the physical sciences receive the lion's share of attention of philosophers of science

have unproblematic counterparts in the real world. We look at two examples that are instructive of the way agents in epistemic logic are different from objects of study in the physical sciences. For instance, there is a difference in modelling the moon and an agent: 'the moon' exists, we could describe and get data about it (and data about its influence on water tides, for example). However, 'The agent' does not exist: models in epistemic logic are always about some hypothetical homogenous agent, not about an actual agent. We want our models to say something about people or other kinds of agents in general, not about 'Sonja Smets' or 'Huub Kasbergen'. We are looking for structures and patterns that transcend an individual actual agent. Now let us compare 'the agent' to some other concept in physics of which there are more; 'the atom' for instance. Atoms seem to follow some well-studied patterns, that we refer to as physical laws. Now, although we may like to refer to rules of our logic as 'laws of thought' (see chapter 3), this does not give them a similar status. People rarely follow these laws.

As we have seen from the few examples, there are many differences between models in the physical sciences and models in epistemic logic, and we should remain aware of such differences. However, this does not mean that the practice of modelling in the physical sciences cannot be instructive for our investigation of modelling in epistemic logic. Modelling in epistemic logic is similar enough to models in the physical sciences in order to apply the DDI account (given in section 4.1.4) to it.

5.1.2 DDI account applied to epistemic logic

It has been argued that we can learn from modelling mostly in the first two stages; the stages of construction and manipulation of the model. In order to see how this works for the practice of epistemic logic, we look at the DDI account of modelling, as presented in chapter 4.

Denotation

The first stage of modelling is the stage of *denotation*, the stage of model construction. This stage consists of several aspects, as discussed by Weisberg.

Assignment.
Firstly, we look at the assignment, which consists of a specification of what is to be studied, following by the coordination of model parts to parts of the target system. In epistemic logic, as we have already seen from the previous section, the object of study can vary quite a bit. However, we can identify the broad lines of the coordination of model parts in standard epistemic logic as in the philosophical and societal agenda.

An epistemic logical model $\mathbf{M} = (W, R_i, \| \cdot \|)$ consists of a non-empty set of possible worlds W, accessibility relation R_i for i in a group of agents G, and valution map $\| \cdot \|$. We call the epistemic logical language $\mathcal{L}$.

The language of our model is related to the natural language. A proposition $\varphi \in \mathcal{L}$ could be any propositional statement like 'it rains' or 'Henry walks the dog'. The possible worlds of the model are related to possible scenarios in the real world: ways the things could have been.[2] The indexes $i, j, k, ...$ stand for several agents. In the real world these could be people, computers and other knowledge-bearing entities. The collection of indexes G, then, stands for a group of those entities. For each agent, we can define an accessibility relation R_i between the worlds of W. This gives, for any possible scenario that the individual may find herself in, the scenarios that she

[2]One could argue about the ontological status of possible worlds, as many philosophers have, but this is outside of the scope of this thesis. An overview of this is given in (Menzel 2016).

considers possible from that point. The accessibility relation reflects the reasoning power of the agent: if the accessibility relation is, say, reflective, symmetric and transitive, we speak about a very powerful agent that among other things is both positive and negative introspective.[3] The valution map $\|\cdot\|$ determines what is the case (which propositions hold) in which situation: it is an auxiliary function to describe possible ways the world could have been.

Now knowledge of some proposition φ, $K\varphi$ or $K_i\varphi$, is defined as having excluded all possible scenarios in which $\neg\varphi$ holds. This is a very strong notion of knowledge defined over our model. As mentioned before, it bears a debatable relation to the notion of knowledge in everyday life.

Besides the model parts of standard epistemic logic, we might find additional or aberrant definitions. For instance, we could add an awareness set, or look at a softer type of knowledge, for which we would add orderings on the possible worlds (for each agent). We will not go into detail about the coordination of all these model parts here, but we may express the hope that every researcher inepistemic logic thinks extensively about the coordination of the model parts in their own system.

Scope.
As has been already extensively discussed, any (epistemic logical) model leaves aside many features of the studied real-world phenomena. Now the scope of our model is meant to determine exactly which aspects of the phenomenon are to be represented by the model and which not. As it is impossible for us to give a general account of the scope of epistemic logical models (this depends on the context of the research question), we may just list some examples of aspects of the real world that are and that are not represented by a standard epistemic logical model.

First we may talk about the agent. What is represented about the agent, is for instance the number of agents, (a very systematic type of) their reasoning power, and the knowledge that the agents have. Besides this, different epistemic logics may add different features about the agent, for example attitudes or awareness. What is not represented about the agent are, for example, limitations in her resources or illogical reasoning, personal details and history, the reasoning process of the agent and embodiment. (cf. 2.4)

Secondly, we look at information. A standard epistemic logical model might capture the propositional statements and informational content. What is not represented in the model, is for instance the mode of representation of this information, and the way the agent gets the information.

Lastly, we look at the scenarios. The model represents at least all the for the question at hand deemed relevant possible scenarios. It is important to note that the scenarios that the model represents are only partial: they only consist of information that is relevant with respect to the current issue.[4] What the model does not represent are all the other things that may go through the human agent's mind. The model does not represent the full, alive, coloured, detailed parts of the scenarios, like images and feelings that are prompted by memories et cetera.

The scope of a model gives us a useful handle to look at the comparison of what *is* represented by the model and what *should be* or *what we want to be* represented by the model. We should carefully note that these things are not the same.

---

[3]Recall that in chapter 2 the relationships between the properties of accessibility relation and the axioms were discussed.

[4]Recall Hintikka's criticism of the term 'possible world'. "In most applications 'possible worlds' are not literally worlds in the sense of universes but merely 'small worlds', that is, so many applications of the language in question, typically applications to some relatively small nook and corner of our four-dimensional world." (Hintikka 2003, 34)

Dynamic fidelity criteria.
The dynamic fidelity criteria describe the error tolerance, determining how well our model should match the real-world phenomena that we are studying. This will be very different for different research agendas. For instance, for a descriptive aim in the societal agenda we may need to get really close to knowledge in the real world, but for solving some philosophical dispute (philosophical agenda) all these details may not matter that much. The dynamic fidelity criteria are related to the assessment of whether our idealisations are acceptable or not.

Representational fidelity criteria.
The representational fidelity criteria determine how well the internal structure of our system should match the internal or causal structure in the target system. It is not easy to come up with criteria for an equivalent of this in epistemic logic, nor is it easy to check whether a model would satisfy those. However, we may remark that the success of possible worlds semantics has been partially due to our intuitive feeling that the internal structure matches our object of study. Fitting said: "A good, intuitive semantics psychologically convinces us we really do understand a subject, correctly or not." According to him, this is why possible worlds semantics changed the field so profoundly: "the right metaphor, the correct intuition, the satisfying feeling." (EL5Q, 86-7)

So even if these criteria are hard to come up with and hard (or impossible) to check, we see that they may make a difference for our practice of epistemic logic nonetheless. If possible worlds semantics had not been so intuitive, would it really have had so much attention?

Demonstration

The second stage of modelling is the stage of demonstration. In this stage the model is analysed, and if needed, adjusted or further refined. We may learn from the refinement or the manipulation of the model. For instance, we may try some situations and see that an agent always has a certain strategy when particular conditions are satisfied. Then we may continue by proving it.

As Hughes mooted, "A mathematical representation should not be thought of simply as an idealization or an abstraction. Like an analogical representation, it presents us with a secondary subject that has, so to speak, a life of its own. In other words, the representation has an *internal dynamic* whose effects we can examine." (Hughes 1997, 330)

The standard epistemic logic has an internal dynamic which can yield us results, theories, insights. As discussed in section 2.5, the various alternatives to standard epistemic logic all have different internal dynamics, thereby yielding us different insights.

Interpretation

The stage of interpretation consists of the assessment of the relationship between the model and the target system, and possibly a translation of the outcomes of the model to the target system.

For instance, when our model brings about certain unintended behaviour of the agent, we might decide it is not yet the right model for our research question, and fix the problem. Obviously we can learn a lot also from such 'failures'. However, the assessment of the model-world relationship may turn out to be more problematic. As has been the critique of Stokhof and Van Lambalgen, we might need a bridge theory. Without such a theory bridging the gap between the formal system in the target system, the results stay within the system and can only take the form of ceteris paribus laws.

The assessment of the model-world relationship naturally depends on the research aim and context. When comparing philosophical arguments (in the philosophical agenda of epistemic logic), for instance, the dynamic and representational fidelity criteria may be quite easy to meet, so the assessment of the model-world relationship will be fine and we can interpret the results achieved within our system without problems. For the societal agenda, for instance, this might be a bit more delicate, and fully depends on the research question.

Learning in the process of modelling

As discussed in chapter 4, it has been said that we can learn from models mainly in the first two stages of modelling. In the first stage, the stage of construction, there are several ways in which we learn. Formalising a phenomenon or concept obliges us to think clearly about all the choices that have to be made in the process of modelling. We have to think about what is derivative and what is basic. We need to think about what we want our axioms to be and why we would like them to be like that. We are forced to make the assumptions of the model precise, teaching us what we tacitly assumed to be true.

In the stage of demonstration, our model may help us learn in various ways (cf. section 2.2, section 3.3.4 and section 4.1.4). Its systematicity may give rigour to what we are trying to find out. Manipulation of a model may give us a new way of looking at things, offering useful metaphors. Sometimes when we manipulate a model and it gives us unexpected behaviour, it may point towards limits of our own human knowledge. As we have seen, also differences between the model and the real world can be very instructive.

5.1.3 Connection model-world

From philosophers of science we learned that abstractions and idealisations are regarded as inherent to modelling. So the formal system always deviates from the target system. We may ask ourselves: what, then, is the relationship between the model and the target system?

Looking back at chapter 4, we may decide to go with Giere's (later) answer: a model is or should be *similar* to the target system in *relevant respects* in a *high enough degree*. Recall that Teller added to this that also those respects in which the formal system and the target system do not agree do matter; we have looked at this in the previous section. However, just as for the natural sciences, we cannot come up with a general account for epistemic logic with respect to the relevance of certain features of the target system. This is dependent on the aim, the intended application and the context of the research question.

Even so, we may advance that it is important to keep realising that our models are not completely faithful to the real world, and often very different from the concept that we are trying to study. Godfrey-Smith noted that, when traditions come into existence in the formalisation process, fictional entities tend to be studied as topics in their own rights. After some time, these fictional entities then are treated as real objects. Exactly the same phenomenon has been observed by Melvin Fitting in the epistemic logical community: "one quickly found oneself talking as if modal notions really were about possible worlds, rather than as if the *formalisation of* modal notions was about possible worlds. It's a distinction to blur." Later, he adds that "it is important that we do not conflate formal methods applied to a field with the field itself. Scientific theories of the worlds are good and useful, but they are not the world itself." The danger of mistaking your formal model for the target system itself exists in both the physical sciences and epistemic logic. However, it should be mentioned that the possibility of an outside check (e.g., to empirical data) reduces this risk. Such an outside check is more anchored in the physical sciences than in epistemic logic.[5]

5.1.4 Changing the question and a bridge theory

Our models are not the world itself.The very process of modelling means changing the question: the phenomenon that is studied is first represented by a model, whose internal dynamics are then used to yield some results. Consequently, the results from this model have to be translated back, in order to say something meaningful about the original object of study. Recall from section 4.2.5 that Stokhof and Van Lambalgen introduced the concept of a 'bridge theory'. When the process of modelling changed the question so thoroughly that the relationship between the phenomenon under study and the resulting model is unclear, an additional theory is needed to demonstrate the appropriateness of the relationship. By means of this concept of the bridge theory, Stokhof and Van Lambalgen meant to distinguish between the practice of the natural sciences and the practice of formal semantics. However, there are also models in the physical sciences that are in need of a bridge theory: those in which de-idealisation of the model could only be derived empirically, not in a theoretical way. Recall from section 4.2.1 that McMullin argued that in such cases the model itself would be "suspect." The process of de-idealisation should be suggested by the idealisations in the model, otherwise they could be considered to be ad hoc.

Now it should be remarked that the need for a bridge theory (in both the physical sciences and epistemic logic) depends on the object of study and the aim of the research. For example, in the formal agenda the objects of study are one or more logical systems. The object of study is not a real-world phenomenon, and therefore no bridge theory is needed: there is no gap between the object of study (the model) and the model. Another example: when the aim is conceptual clarification, as is sometimes the case in the philosophical agenda of epistemic logic, there is not always the need for a bridge theory. One can use the model instrumentally, while being aware of its limitations and its unclear relationship to the target phenomenon. On the other hand, when the aim is descriptive and we intend to let the model say something about the real-world, as can be the case in the societal agenda of epistemic logic, the critique of Stokhof and Van Lambalgen may be seen to hold for epistemic logic as well.

When the relationship of a formal system is assessed to be inadequate, if it needs a bridge theory that cannot be provided, then the results achieved within the formal system cannot be automatically extrapolated to the target system. Such results may only take the form of ceteris paribus laws, like 'X is the case under the ideal circumstances that I'. As long as the bridge theory is not provided, the extent of the use of our models remains unclear. Again we should remark that the assessment of the model-world relationship depends on the research context. Therefore, we will look at possible research contexts in epistemic logic and their connection to the acceptability of idealisations more closely in the next section.

5.2 Abstractions and idealisations in epistemic logic

Throughout this thesis we have often concluded that we cannot give a general answer of the acceptability of abstractions and idealisations unless we look at them in their research context. There always needs to be a balance between having a clean model that is easy to use and having a faithful representation of the target system. Now for one side of this balance we need to look at what the model is used for, in order to establish which properties make the model useful. For the other side of the balance we also need to know the aim or intended applications of

[5]However, not all formal models used in the physical sciences have an unproblematic outside check.

the model, in order to establish to what extent and in which aspects the model is supposed to stay faithful to the target system. We need to know what the model is supposed to do, if we want to evaluate whether the abstractions and idealisations are considered to be a problem or not.

In this section we will have a more detailed look at what kind of research contexts we may find in the field of epistemic logic. The examples are instructive of what kind of considerations may play a role in evaluating the abstractions and idealisations of the model. Furthermore, we will have a look at a few more issues that have come up when discussing models of epistemic logic.

5.2.1 Abstractions and idealisations in various research agendas

Recall from chapter 3 that a categorisation of the field of epistemic logic got suggested, in order to get some grip on the diversity of epistemic logical research context and opinions of the researchers. We will discuss the four research agendas separately.

Formal agenda

As the object of study of the formal agenda of epistemic logic is not any real-world phenomenon, but rather the logical systems that may be given an epistemic interpretation, abstractions and idealisations do not constitute a problem towards this agenda. Indeed, there are no abstractions and idealisations at all in the formal agenda, as the formal system does not deviate from the target system (the formal system is the target system). Therefore, the formal agenda suffers the least from possible attacks along the lines of not being realistic enough.

Of course, the motivation for mathematically studying the formal system, may come from one of the other agendas, as often is the case. To those aims abstractions and idealisations may constitute a problem, but not to the formal work. For example, if we prove that neighbourhood semantics has the property that it's closed under logical equivalence, this may be problematic if we intended to solve the problem of logical omniscience with it in order to model human knowledge.[6] However, this is still an unproblematic result for our formal agenda.

Philosophical agenda

To the aim of conceptual clarification, idealisations do not necessarily and do often not constitute a problem. As we have seen from 3.3*vi*, we have many possibilities of learning from idealisations. In some cases, it might be even better to consider ideal situations, and going beyond our subject material may give us a new look at its limitations.

In the philosophical agenda we often speak of the 'rational man' (see 3.3.*viii*). When we do this, we change the object of study and therefore our evaluation of possible abstractions and (false) idealisations change. If the object of study is human knowledge, then logical omniscience is an idealisation (when mentioned in the discussion of the model, otherwise it may be a false idealisation). If the object of study, however, is postulated to be the rational man, then logical omniscience is not a deviation from the target system, and thereby neither an abstraction nor an idealisation. Nonetheless, the price we pay for this move is high: we are no longer talking about real agents.

For our rational construction to give a norm for human reasoning or behaviour, we need the model to be close enough to this target system (see also 3.3.*iii*). Otherwise it may be unclear how to even try to live up to the norm. Besides, it should be mentioned that classical logic is not

[6] Although one could argue that also such 'failures' in science, when something does not work as you expected it to work, are very valuable indeed: finding out ways that do not work also bring about scientific progress.

always the best logic to reason with. It can be advantageous to jump to conclusions. Imagine, for instance, that you're deep in a forest and you suddenly see a bear in front of you. In that case you better give the appropriate bear-response (if the bear is not attacking: slowly retreat), instead of contemplating about the possibility of hallucinating or being a brain in a vat that sees the bear because of some electronic stimulus, and so on. So also whether you can and whether it is wise to live up to a norm of rationality or not, depends on the context.

Societal agenda

With the descriptive aim of the societal agenda of epistemic logic, we ultimately want to model knowledge realistically. The societal agenda in this way is the only research agenda of epistemic logic that has 'de-idealisation' (see chapter 4) as one of its main aims. To that extent, the societal agenda suffers the most from the idealisations inherent to epistemic logic. However, whether the idealisations are conceived of as problematic or not, depends on the individual researcher too: we may also look at it pragmatically. Not idealising does not get is anywhere either (3.3.*i*) and we may find in the ideal models a starting point (3.3.*ii*) for further refinements. That is, we may temporarily tolerate the idealisations in hope of finding a way out of them at some point in time.

Because the societal agenda has most to do with abstractions and idealisations, we have a look at how our new definitions of abstraction, idealisation and false idealisation work. Suppose some researcher modelled how information proliferates among scientists. In the discussion accompanying the model some of the relevant features that the model does not capture realistically are mentioned, for instance, (lack of) logical omniscience and the mode of presentation of the information. These are the model's idealisations. Its abstractions are the irrelevant features of the target system that the model does not represent and that also the discussion stays silent about: the hair colour of the researchers, their nationalities, whether they have pet animals, et cetera. Now another researcher may come with the results of a psychological study showing that people (also scientists) suffer from the cognitive bias that has been called the 'halo effect': the overall (positive) impression we have got of a person (positively) influences our evaluation of that person's specific traits. For example, if we like someone, we are more likely to think they are smart, we are more likely to trust their words. Likewise, if we do not like a person, we are less likely to think they are smart, we are less likely to trust their words (this negative cognitive bias goes by the name 'horns effect'). So, by bringing up this psychological study, the second researcher may suggest that actually scientists' attitudes towards each other are relevant to the information proliferation between them, and should therefore be represented in the model or mentioned in model discussion. The second researcher so claims that the model, that was presented to be a idealised model, actually contained a false idealisation: it silently omitted a relevant feature. Depending on the way the model has been construed, the first researcher may choose to either incorporate this feature in the original model or abandon the model and look for a better one (or, of course, defend the original model).

Although the descriptive aim of the societal agenda may suffer the most from idealisations, the other side of the coin is that researchers with a societal agenda have most of all a possibility of an outside check[7]. This can be to empirical data, but also to the opinions of researchers in other fields that they collaborate with. Therefore, they have less of a danger to get into an ivory tower situation.

Sometimes in the societal agenda one may have a normative aim, for example, the aim of avoiding some cascade. In this case the same comments hold as in the normative part of the philosophical agenda. If we find a way out of some informational cascade, it would not be very useful if we

[7]However, as mentioned in section 3.3.1, gathering and interpreting empirical data is not always on problematic.

would say: 'if everyone from now on just follows through all the consequences of their knowledge, we will be just fine.' We need the norms that we want people to live up to be realistic enough in order for us to be able to (try to) live up to them.

Computer science agenda

The computer science agenda is the most pragmatic of all. in this agenda we usually have a particular intended application in mind, which enables us to evaluate the model with respect to its usefulness. Therefore the modeller him or herself can best assess whether the system does what it is supposed to do. For example, if the intended application is an interpreted system, then the **S5**-assumptions are assessed to be appropriate (see the discussion of axiom (5) in 2.3.3). However, the intended application could also be, for instance, a machine that reasons like a human being. In this case the idealisations do constitute a problem towards the aim, and we would likely have our research project crossing over to the societal agenda too.

5.2.2 The problem of logical omniscience revisited

Logical omniscience is quite a special kind of idealisation, as it is not introduced consciously, but rather appears as a byproduct of the chosen formalisation. Recall from section 4.2 that the philosophers of science almost unanimously spoke of idealisations as *deliberate* deviations from the reality of the target system. However, logical omniscience is not a deliberate deviation from human knowledge, and it is seen (in section 2.5) to be very hard to escape. Although the alternative approaches appear to be different, they have quite some similarities in their structure. They complement or replace the Hintikka/Kripke definition of knowledge (as truth among all accessible worlds) by some other condition that tells us when some sentence or proposition can be known. The impossible worlds semantics are most original in this sense: they change the set of worlds that the agent considers, instead of just changing the definition of knowledge.

However, most importantly, in all solutions we see the tension between our desire to equip our agents with logical abilities or 'rationality', and our desire to stay close to an everyday conception of knowledge. Human reasoning is not as consistent as the logic underlying our frameworks, and often depends on many parameters. Logical omniscience is the price we pay for wanting to model logical abilities of the agent in a systematic way. As Stalnaker remarked, logical omniscience is a "conceptual problem we all face, and not one that is subject to a technical fix." (EL5Q, 247) Indeed, the (technical) alternatives raised so far have not been very fruitful as a general solution to the problem of logical omniscience. Our wish to construe our agents as rational beings is at the root of the problem, consequently all paradigms of rationality face the problem of logical omniscience.

Just like logical omniscience is a hard idealisation to escape for epistemic logic, the same holds for other methods of modelling knowledge that define their agents to be rational. We can compare our discussion of dealing with logical omniscience in epistemic logic with how quantitative methods of modelling knowledge deal with logical omniscience. The solutions raised to the logical omniscience problem in epistemic logic are many more in number and diverge more from the original framework (standard epistemic logic) than the quantitative approaches to modelling knowledge. A look at the literature[8] shows that in the latter one rather tries to avoid the problem of logical omniscience indirectly, while keeping the whole mathematical machinery of probability intact. A reason for this could be that in the quantitative approaches to modelling knowledge, logical omniscience is not perceived as a main problem, but rather discussed as a side issue in connection to other issues. Examples of such connected issues in the quantitative approaches to

[8] E.g., De Finetti 1972, Gaifman 2004, Garber 1983, Lehman 1955.

modelling knowledge are the Dutch book argument (De Finetti 1972) and the problem of old evidence (Garber 1983). A possible reason for keeping the core of probability theory intact while trying to model more realistic reasoning, could be that probability theory emerges from a long philosophical discussion on the interpretation of probability itself. Although the birth of epistemic logic also has known a lot of philosophical discussion, the systems of epistemic logic did not emerge from it. Rather, a nice mathematical framework existed, and the suggestion to interpret modal logic epistemically gave birth to the field of epistemic logic.

Although other approaches to modelling knowledge may also include idealisations like logical omniscience, their way of treating them can be completely different. In epistemic logic, logical omniscience is often considered to be a problem that has to be solved. As we saw from section 2.5, sofar the 'best' solution remains to be application-dependent. This is not necessarily a problem; in section 5.4 we will argue for a multiple-model approach: we do not always need all our aims to be achieved within one model. However, researchers are looking for a more satisfying solution to the problem of logical omniscience; they are looking for models that are more faithful to knowledge as humans have it. In the next section we discuss prospects of more realistic logical modelling.

5.2.3 Need for more realistic models

From the societal agenda, but also from other areas we have (in chapter 2 and 3) heard the call for more realistic models. This can be, for example, in order to understand bounded reasoners better, to build human-like machines, to make epistemic logic suitable for applications in cryptography, to model players in epistemic game theory more realistically, et cetera. We may conclude that there are applications for which we are in need of more realistic models than those based on standard epistemic logic.

However, we also heard the warning by Scott and others, that we should make sure that we are not just refining our models beyond the level of applicability. First we need to understand our concepts and our formal systems well enough. Whether or not epistemic logic has indeed refined its models beyond the level of applicability, will have to be a topic of further work.

Underlying logic is classical What most epistemic logical systems have in common, even the alternatives raised as a solution to the problem of logical omniscience, is that their possible worlds are defined as maximally consistent sets of propositions of the language. The world is being maximal implies that they cannot be incomplete, the worlds being consistent implies that it can be inconsistent. The language underlying standard epistemic logic and most of its discussed alternatives is classical propositional logic. Impossible worlds semantics do give us the opportunity to have incomplete and inconsistent worlds. Other approaches that try to avoid logical omniscience set extra requirements for knowledge, to make it a bit more realistic. But perhaps there are other logics that would more suitably underlie our framework. Perhaps, for some applications, we need a more drastic solution.

Looking for a better formalism As we saw from section 3.4, there are many reasons why we could consider the idealisations of epistemic logic to be acceptable. However, the epistemic logics we investigated are not appropriate for all applications, and not everyone is completely happy with using modal logics for modelling epistemic situations. Fitting mentions that "nothing rules out the creation of a better formalisation based on different insights." (EL5Q, 86) Indeed, some may even be looking for such a creation. As we may recall, Bílková encourages us to be creative in this process of looking for a better formalism. We should use all the mathematics that we have available to us, start anew and be open-minded about a possible new formalism for epistemic interpretations. In other words, we should rethink using modal logics.

5.3 Being explicit

Browsing through the epistemic logical literature shows us that aims, claims and success criteria of epistemic logic are sparsely explicitly discussed. Although researchers may think about these things and discuss them informally together, seldom these thoughts and discussions get written down. However, among the common heading of the field of epistemic logic, there are different practices and accordingly different methodologies to be distinguished. The introduced categorisation in four agendas meant to attract attention to this fact.

At the same time, when assessing the relation of your formal system to its target system, it is important to know what is your research context. Depending on the context, some feature may be deemed relevant or irrelevant and consequently, it may count as either an abstraction, an idealisation or a false idealisation. Depending on the context, some idealisation may be seen to either constitute a problem to the research aim are not.

Therefore, it is important that people are aware of the research context, this should be made explicit instead of implicitly assumed to be clear. Being explicit about the idealisations too brings them up for discussion, hence their evaluation can be more critical, open and complete. There are two reasons why this is important: (1) it gives the researchers a more faithful self-image, and (2) it enhances communication.

5.3.1 A more accurate academic self-image

As has been discussed previously (summarised in section 5.1.4), although we cannot come up with a general account of the relationship between the formal system and its target system, the formal system is often very different from the target system. Both Godfrey-Smith (philosopher of science) and Fitting (epistemic logician) warned us not to confuse our models with the reality of their target systems. Both in science generally, and epistemic logic in particular, the distinction between our formal notions and reality seem to at times have been blurred.

As discussed in the first section of this chapter, the extent to which the results reached in an epistemic logical system can be extrapolated to its target system are not always clear. Sometimes we need an additional theory to bridge the gap between the formal system and the real-world phenomenon or concept it is trying to model. Where such a bridge theory is needed and lacking, we cannot be sure that the outcomes of the formal system hold for the target system. This is an extra reason why we should be clear about our aims and the idealisations of our models: we need to be clear about the extent to which they say something about the target system. If we do not, we may (implicitly) take them to do more than they actually do.

5.3.2 Better communication stimulates interdisciplinarity

Not only among ourselves, but also towards people outside our own research fields, we find a reason to be clear about our idealisations and research contexts.

It has often been mentioned that one of the merits of epistemic logic is its interdisciplinarity. Van Benthem encouraged us to think in themes instead of subdisciplines (section 2.3). However, for such an approach to research we need to make sure that the notions that we use and our methodology are clear for others to understand and relate to. We saw from the introduction of section 2.3 that often different words are used to refer to the same or similar concepts. Likewise, we saw from section 4.2, that in the scientific community also the same words are used to refer to different concepts. And not only that: also notations are seen to differ greatly, hampering communication even more. The notations of the different systems that were introduced in section

2.5 (for didactic purposes they have been discussed in a similar make up), differ a lot in their original forms.[9] Besides a different use of words and notation, research aims and contexts often remain implicit.

To the researchers from other fields we collaborate with, but also to 'real' outsiders we want to be clear about our research aims, methods and considerations. In order to truly understand the merits and subtleties of the work done in a particular field of science, it seems that one needs to be working in this field. However, it would be helpful for the communication towards the outside world if we stated more explicitly what our models are meant to do, what they do right and what could be improved. Perhaps some of the critiques that epistemic logic received[10] would have been milder, if one had better understood the research aims and methods.

Besides, Löwe warned us for the additional risk of getting into an ivory tower situation. Better communication to the outside world could open up the research field, thereby stimulating a more accurate academic self-image (5.3.1).

Discussion In science generally, research contexts often remain implicit and the notions that we use vague (recall the different uses of the words 'abstraction' and 'idealisation' in philosophy of science, discussed in chapter 4), this is however no reason for epistemic logic to do the same. When asked about the merits of using (epistemic) logic, logicians proudly refer to how we make tacit assumptions explicit and up for evaluation. Exactly the same we should do for our own assumptions: make them explicit so that they are up for discussion. This may not be enough, and over time we may still find the need to construct more realistic models for certain purposes, but it is a start.

5.4 Multiple models

Tying to our discussion of wanting to be interdisciplinary, we may have a look at multiple ways of modelling knowledge and knowledge transfer. Recall from chapter 4 that Weisberg mooted that there are three types of idealisations, reflecting three different practices in science. The third type of idealisation encompassed constructing multiple models that together add up to a fuller comprehension of the target system. Weisberg's 'multiple models idealisation' is best suited for a type of research practice that has a complex aim. For example, the aim of conceptual clarification, or more generally: the aim of scientific progress. These aims suit the practice of multiple models very well: different types of models may provide us with different angles so that we get a diverse and broad understanding of the concept. The results of the different types of models add up so that we may get closer to a complete picture.

So, rather than looking for the one perfect system to capture knowledge, we may agree that different models together contribute to a better comprehension of this complex and diverse concept. We may combine insights from different epistemic logical frameworks. Indeed, we may, as Bílková suggested, be mathematically creative and look for different systems than modal logic to be interpreted epistemically. Moreover, in order to get a more complete understanding of the concept of knowledge should not only use logic, also quantitative methods and other fields of science may contribute to this. As Hansson phrases it: "We need a variety of models of epistemic processes. Logic does not have a monopoly." (EL5Q, 130)

[9]To get an impression of this, have a look at, for example: Fagin et al. 1995, Artemov and Fitting 2006, and Halpern and Pucella 2011.

[10]Recall Castañeda's review of Hintikka's book *'Knowledge and Belief'*.

Then again, in order for researchers from different fields to understand each other and the able to profit from each other's insights, we again come back to the need of being explicit about our research practice.

5.5 Conclusion

Idealisations are inherent to modelling: instead of studying a real-world phenomenon or concept directly, we change the question and study a formal system. The formal system has internal dynamics that we can learn from. In order to say something about the original object of study, we then need to translate the results of the formal system back to results about the target system. This interpretation, however, is not always unproblematic. Sometimes the model-world relationship is unclear and we need an additional theory that bridges the gap between the formal system and the target system.

The formal system is assessed to have the appropriate relationship to the target system if it is similar to the target system in relevant respects and in a high enough degree. Which aspects of the target system are relevant, and what counts as a high enough degree, depends on the context.

For the field of epistemic logic the research contexts and research aims may differ from project to project. Some epistemic logicians concentrate on formal work, others more on the philosophical prospects that are opened up. For some researchers epistemic logic is more normative, for others it is more descriptive. Some epistemic logicians relate to research in the social sciences, and others use epistemic logic pragmatically to study networks or computer programs.

Depending on your research agenda, idealisations may more or may less constitute a problem towards your research aim. In general, there are many ways in which we can learn from and despite the idealisations we make. But we need to find the right balance between nice mathematics and a realistic representation of the target system, a balance that depends on what your model is supposed to do. For many applications, idealisations are unproblematic, but there are also applications that ask for more realistic models. As there is always the possibility of the existence of a better formalism, we may be mathematically creative and look for systems not based on possible worlds that can be given an epistemic interpretation.

Whatever the goal, being explicit about the aims we have and the idealisations we make is essential. At the moment, in epistemic logical literature the aims, success criteria and overall research contexts are sparsely written down and therefore often remain implicit. If we were to be more explicit about those considerations, it would benefit the field of epistemic logic in at least two ways. Being explicit may cause our academic self-image to remain appropriate by pointing to the extent to which and for what exactly we can use our models. This enables us to distinguish between models that suit our intended applications well, and models that do not. Furthermore, being explicit improves communication, both within and outside the research community. Better communication between epistemic logicians and researchers in other fields enhances interdisciplinary research. This is wanted, because multiple ways of modelling yield a fuller comprehension of our object of study: knowledge.

5.6 Suggestions for further work

Throughout this thesis project, interesting questions came up. Because we could not discuss or answer all questions, here are a few interesting questions to ponder on:

- Perhaps someday someone comes up with a logic that captures the (apparent) whimsicalness of human reasoning. Will we then still call it logic?
- To what extent do we want to refine our epistemic logical models to suit more realistic situations? Is there a point at which bringing yet another logic for yet another particular situation is not really contributing any more to the progress of the field of epistemic logic? If so, have we already passed this point?
- Are we refining our models beyond the level of applicability?[11]
- Questions of success: Who is to judge whether a formal framework is successful or not? What does it mean for a logical framework to be successful? Are there any necessary or sufficient conditions that hold in general? Is failure possible, and if so, when is an epistemic logical work a failure? As universities do not have infinite financial resources, we could tie questions of success to questions about funding: Which considerations play a (decisive) role in allocations of funding, and do these considerations really contribute to the 'success' of the epistemic logical community? In funding, do and should we strive for the success of the university, or of the whole research field?

Furthermore, there have been some research directions that I would like to have explored further, but that I could not, because of limitations in time (and thesis length).

- Epistemic logic is often said to be useful for many things, see section 3.3.4 for an overview of suggestions from the interviews and the literature. However, this usually does not become concrete. For example, some epistemic logicians see a role for epistemic logic in aiding empirical research. However, how exactly epistemic logic helps, and which research projects are examples of this, remains unclear. An interesting research direction would be looking for such examples and evaluating if and how epistemic logic plays a role here. More broadly, it would be good to investigate and be really explicit about in which cases epistemic logic has been useful, and how. This investigation should also look at the bigger picture: why is it important that we solve the formal issues that we solve? Having a formalisation of the muddy children puzzle is nice, but is there a further application or goal on the horizon that we need it for?
- It would be interesting to have an extensive look at the epistemic logical literature to check which claims are made, and if the work that makes the claims also proves them or shows them to be feasible. One could, for instance, imagine an epistemic logical article that has a philosophical claim in the abstract and introduction, but further only consists of formal work. More broadly, an interesting research direction would be to investigate to what extent epistemic logicians possibly confuse their models with the target systems (as suggested by Godfrey-Smith and Fitting in sections 4.1.4 and 5.1.4).
- As discussed at the beginning of chapter 3, research aims and contexts often remain implicit in epistemic logical literature. It would be interesting to investigate if this is different for introductory textbooks or textbook chapters on epistemic logic. One could imagine that idealisations, research aims, and intended applications are discussed more elaborately in such texts, as they are written for laymen.
- As a follow-up study on section 5.3, one could investigate how disparate various fields of science that study knowledge are, and if scientific progress would be stimulated if there was more interdisciplinary research. Furthermore, it would be interesting to see to what extent better communication (in particular, being explicit about the research context and idealisations) could contribute to that.

[11]Recall from section 3.4.*iv* that Scott raised this point.

Epilogue

I learned a lot from writing this thesis. Not only about epistemic logic and its practice, but also more general lessons that are valuable in the process of writing a thesis, and even beyond that.

Go beyond your subject material Like biologists sometimes study three-sexed organisms, and epistemic logicians study rational agents with infinite resources, I too went beyond my subject material. I investigated the practice of modelling, and in particular its abstractions and idealisations, in the sciences generally: looking at examples and literature from physics, economics and biology.

Idealising is useful Just like idealisations in epistemic logic often get accepted because of their usefulness (see section 3.3), I too had to narrow down the complexity of my subject material by introducing idealisations. In chapter 3 I moot a categorisation of the field of epistemic logic into four research agendas. Of course, the field of epistemic logic does not fall into four parts that are mutually exclusive and collectively exhaustive, nonetheless it is very useful to discuss them separately.

It's all about balance Just like in our epistemic logical models we seek for a balance between the mathematical and the real based on our research aim and context, in this thesis I had to seek for several balances. On the one hand, I want to present a complete picture, on the other hand, I want a structure that is not too complex, in order for us to oversee, explain and understand the subject material. Likewise, in my methodology and throughout the chapters when I am motivating choices that are made, I want to both give a complete justification as well as limit the length of this thesis.

Be explicit about the assumptions you make Just like epistemic logic is praised for making tacit conditions, underlying assumptions and regulations explicit, we find the need for epistemic logicians to be explicit about the assumptions they make (see chapter 5). This is also what I tried to do while writing thesis: I tried to be as explicit as possible about, on one hand, all the idealisations that I see, and on the other hand, on a meta level, about the simplifications I make.

Multiple approaches add up to a better understanding Like Weisberg argued for the sciences, and some epistemic logicians (among whom I mention Van Benthem, Fitting and Hendricks in this thesis) for the field of epistemic logic, sometimes multiple approaches add up to a fuller comprehension of the subject matter. Likewise, I have been diverse in my approach to the subject material of this thesis: looking at the literature of epistemic logic, conducting interviews with researchers in the field of epistemic logic and comparing these findings with findings from the literature of the philosophy of science. On another scale this lesson about the value of multiple approaches holds too: although I suggest a categorisation of the field of epistemic logic, I am aware of the possibility that there are different approaches here. I am open to discuss them when raised, as I am of the opinion that this will likely bring some progress in the new terrain of philosophy of epistemic logic.

Appendix A

This appendix gives an introduction to modal logic.

Modal systems

Modal logics can be seen as a family of logics each of which give a different interpretation of the underlying structure. Syntactically, the modal logical language is a combination of the propositional logical language and some modal operator '$\Box_i$', referred to as 'box'. Modal logic is called 'modal' because its language can express a distinct modality, e.g. a qualification of the truth, of sentences. The modal auxiliaries in English include can, could, may, must, need, ought, shall, should, will, and would.[12] The initial interpretation for this modal operator was in terms of necessity: '$\Box_i\varphi$' can be read as 'it is necessary that φ'. Later interpretations have been given in terms of, amongst others, obligation, time, knowledge and belief.

Formally, the modal logical language $\mathcal{L}_{\mathcal{ML}}$ is defined as follows. To be general, we define it with respect to multiple agents immediately; for the single agent case we just have to drop the indexes.

Definition 5.1. [13] *(Modal Logic language $\mathcal{L}_{\mathcal{ML}}$). For G a set of agents, $i \in G$ and At a basic set of propositions (atoms) named by $p, q, r, \ldots$, we may inductively define how to construct further expressions:*
$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box_i\varphi$

This means that any formula in this language is constructed from atoms, negations, conjunction and the box operator. Note that $\varphi \vee \psi$ and $\varphi \rightarrow \psi$ can be constructed by using negation and conjunction.[14] The truth conditions for the last concept, $\Box_i\varphi$, will be given shortly after.

In modal semantics, the aforementioned possible worlds and a binary relation between those worlds are introduced. Possible worlds are taken to be maximally consistent sets of propositions. The binary accessibility relation, often indexed with i for some agent i, determines which worlds agent i has access to from her point of view (whatever that point of view may be). That is, which scenarios the agent considers possible viewed from a certain (actual) world. The features of the accessibility relation, e.g. symmetry or transitivity, determine how much access an agent has to the set of possible worlds. A frame $F =< W, R_i >$ consists of a set of possible worlds W and such an accessibility relation R_i, and is –as we will see– an important notion in modal logic. A possible worlds model, also referred to as Kripke-model, is a frame with some valuation $\| \cdot \|$ that assigns a set of worlds $\|p\|_M \subseteq W$ to each proposition signifying the set of worlds in which proposition p is true. A model with a specifically assigned 'actual world', (M, w), is called a pointed model, and often the parentheses are dropped. Formally, all this is defined as follows.

[12] Source: the free dictionary, thefreedictionary.com.
[13] This definition follows Van Benthem 2010.
[14] $\varphi \vee \psi \equiv \neg(\neg\varphi \wedge \neg\psi)$ and then $\varphi \rightarrow \psi \equiv \neg\varphi \vee \psi$.

Definition 5.2. *(Possible Worlds Model)*[15][16]*. For At the set of propositional atoms and G the set of agents, a possible worlds model for agents* $i \in G$ *is a triple* $\mathbf{M} = (W, R_i, \| \cdot \|)$ *such that*
(i) W *is a non-empty set of possible worlds;*
(ii) R_i *is an accessibility relation between the worlds of* W*;*
(iii) the valution map $\| \cdot \|$ *assigns to each atomic sentence* $p \in At$ *some set* $\|p\|_M \subseteq W$ *of worlds in which* p *holds.*

The truth conditions in the possible worlds model are as follows, whereby $\Box_i$ is defined as truth over all accessible possible worlds. $M, w \models \varphi$ means that φ is true in model M at world w.

Definition 5.3. *(Truth conditions)*[17]*. Every possible worlds model M and world w satisfy:*
$M, w \models p \Leftrightarrow w \in \|p\|_M$
$M, w \models \neg\varphi \Leftrightarrow$ *not* $M, w \models \varphi$
$M, w \models \varphi \wedge \psi \Leftrightarrow M, w \models \varphi$ *and* $M, w \models \psi$
$M, w \models \Box_i\varphi \Leftrightarrow$ *for all* w' *with* wR_iw'*:* $M, w' \models \varphi$

Modal logic is useful and interesting for many reasons. For one, the approach is local, which enables considered possibilities of an agent to be sensitive to her point of view. Furthermore, the most interesting modal axioms all correspond to some condition on frames. These axioms and corresponding frame conditions make up for different modal logical systems. We will have a look at these different systems now.

The simplest, minimal system in the Kripkean structure defined above, is system **K** (for Kripke). It satisfies the distribution axiom and necessity. The distribution axiom (K) is defined as follows:[18]

(K) $\Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$.

We can check that this is satisfied in any possible worlds model:
Whenever in an arbitrary model M and arbitrary world w it is the case that $M, w \models \Box_i(\varphi \to \psi)$, this means that for all w' with wR_iw': $M, w' \models (\varphi \to \psi)$. So in each accessible world either φ is false or ψ is true. So then, if $M, w \models \Box_i\varphi$, it follows that $M, w \models \Box_i\psi$. This shows that the distribution axiom holds in any (normal) possible worlds model.

Next to consider is the inference rule of necessitation (N):

(N) from φ infer $\Box_i\varphi$,

stating that every proposition that is valid, is necessary.[19] It can be seen from the definition of $\Box_i$ that this too holds in minimal modal system **K**.

More interesting for actual use, however, are the stronger systems that we can create by adding axioms to minimal system **K**. We will discuss a few of them, the first being system **T**, that is constructed by adding to **K** the axiom (T):

[15] Definition following Baltag and Smets 2010.
[16] In line with common practice in epistemic logic, we let W be countable, G finite and At infinite.
[17] This and rest of the definitions in this section follow Van Benthem 2010.
[18] It is a convention that each axiom should be understood universally. Indexed or not, the distribution axiom, for instance, may be read as: 'for any agent i, any formulae φ and ψ, in any model M and world w, it is the case that $M, w \models \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$.
[19] Of course, this inference rule should be carefully distinguished from the statement that 'if φ is *true*, it is necessary', which is clearly not valid. However these two are often confused among (new) students of modal logic.

(T) $\Box_i\varphi \to \varphi$,

stating that whatever is necessary is the case. This corresponds to a reflexive binary relation, so **T** corresponds to the class of reflexive frames.

From system **T** we can construct system **S4** by adding axiom (4):

(4) $\Box_i\varphi \to \Box_i\Box_i\varphi$,

stating that whatever is necessary is necessarily necessary. This corresponds to a transitive accesibility relation, so **S4** corresponds to the class of reflexive and transitive frames.

Even stronger, we can add axiom (5) to **S4**, giving us system **S5**. Axiom (5),

(5) $\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi$,

is associated with euclidean[20] accesibility relations. The system system **S5** corresponds to the class of reflexive, symmetric and transitive frames.[21]

The discussed axioms and corresponding conditions on frames are given in table A.1. All the discussed logics are so-called 'normal modal logics': modal logics that are extensions of system **K**, satisfying both axiom (K) and inference rule (N).

Table A.1: Axioms and corresponding frame conditions (based on Garson 2016)

Name	**Axiom**	**Frame condition**: $\forall u, v, w \in W$:	**R is...**
(K)	$\Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$	none	no specification
(T)	$\Box_i\varphi \to \varphi$	wRw	reflexive
(4)	$\Box_i\varphi \to \Box_i\Box_i\psi$	if wRv and vRu, then wRu	transitive
(5)	$\neg\Box_i\varphi \to \Box_i\neg\Box_i\varphi$	if wRv and wRu, then vRu	euclidean

[20]See table A.1.

[21]It can be shown that reflexive, transitive and euclidean binary relations are symmetric.

Appendix B

This appendix provides all the formal definitions and technical details needed for understanding section 2.5 on logical omniscience.

Standard epistemic logic

Standard epistemic logic suffers from all seven forms of logical omniscience, as introduced in section 2.5. LO(1) is equivalent to inference rule (N) (discussed in Appendix A), LO(4) is equivalent to axiom (K). Also the other forms of logical omniscience can be checked to hold in any standard epistemic logic. We look at two instructive examples.

Let us first have a look at closure under logical implication. Suppose $\vdash \varphi \to \psi$. Then by inference rule (N) or 'closure under theoremhood' it follows that $\vdash K(\varphi \to \psi)$. Now by axiom (K) or 'closure under material implication' it follows that $\vdash K\varphi \to K\psi$. So we see that closure under logical implication, LO(2), follows from closure under theoremhood, LO(1), and closure under material implication, LO(4).

We look at closure under conjunction. Suppose for contradiction that a normal modal logic is not closed under conjunction. So we can find a possible worlds model $\mathbf{M} = (W, R, \| \cdot \|)$ and a world w so that the following holds: $M, w \models (K\varphi \wedge K\psi)$ but $M, w \not\models K(\varphi \wedge \psi)$.
Now, by definition 2.4 of knowledge K, it holds that $M, w \models (K\varphi \wedge K\psi)$ iff $\forall w' \in W : wRw'$ it holds that $M, w' \models \varphi$ and $\forall w'' \in W : wRw''$ it holds that $M, w'' \models \psi$. Now, again by definition of knowledge K, it holds that $M, w \not\models K(\varphi \wedge \psi)$ iff $\exists x \in W : wRx$ and $M, x \not\models (\varphi \wedge \psi)$. This holds iff $M, x \models \neg(\varphi \wedge \psi)$ iff $M, x \models \neg\varphi \vee \neg\psi$, which holds iff $M, x \models \neg\varphi$ or $M, x \models \neg\psi$. Now in the first case we reach a contradiction with $\forall w' \in W : wRw'$ it holds that $M, w' \models \varphi$, and in the second case we reach a contradiction with $\forall w'' \in W : wRw''$ it holds that $M, w'' \models \psi$. We may conclude that in normal modal logics it is not possible that $M, w \models (K\varphi \wedge K\psi)$ while $M, w \not\models K(\varphi \wedge \psi)$, so normal modal logics are closed under conjunction, LO(6).

Syntactic logic

Definition 5.4. *For model* $\mathbf{M} = (W, \| \cdot \|)$*,* $\mathcal{L}$ *the language based on the set of atoms At, and syntactic knowledge function*[22] $\mathcal{C} : W \mapsto \mathcal{P}(\mathcal{L})$*;*

$M, w \models K\varphi$ *iff* $\varphi \in \mathcal{C}(w)$

[22] This definition follows Halpern and Pucella 2007.

Awareness logic

Definition 5.5. *For epistemic possible worlds model* $\mathbf{M} = (W, R, \mathcal{A}, \| \cdot \|)$*,* $\mathcal{L}$ *the language based on the set of atoms At, and awareness function*[23] $\mathcal{A} : W \mapsto \mathcal{P}(\mathcal{L})$*;*

$M, w \models K\varphi$ *iff* $M, w' \models \varphi$ *for all* $w' \in W : wRw'$ *and* $\varphi \in \mathcal{A}(w)$

Justification logic

Let $\mathcal{JL}$ be the language of the justification logics.

Definition 5.6. *(Justification logical language* $\mathcal{JL}$ *)*[24] *The language* $\mathcal{JL}$*, can be inductively defined:*
$t ::= x \mid e \mid (t \cdot t) \mid (t + t) \mid !t$,
$\varphi ::= p \mid \bot \mid (\varphi \to \varphi) \mid t : \varphi$

This means that each justification term t can be build up by justification variables $(x, y, z, ...)$, evidence or justification constants $(e_1, e_2, ...)$ in combination with the operators 'application' $(\cdot)$, 'sum' $(+)$ and 'internalisation' (!). Each formula φ is built up just like in standard epistemic logic. Statement $t : \varphi$ can be read as follows: 'justification term t serves as a justification (or proof) of formula φ.

Truth in a possible worlds justification model is similar to truth in modal logic, except for the definition of knowledge (or being justified in believing):

Definition 5.7. *For possible worlds justification model* $\mathbf{M} = (W, R, \mathcal{E}, \| \cdot \|)$*,* $\mathcal{JL}$ *the justification logical language, and evidence function* $\mathcal{E} : T \times \mathcal{JL} \mapsto \mathcal{P}(W)$*;*

$M, w \models (t : \varphi)$ *iff* $M, w' \models \varphi$ *for all* $w' \in W : wRw'$ *and* $w \in \mathcal{E}(t, \varphi)$

The evidence function maps combinations of terms and formulae to set of possible worlds. The idea here is that if for some possible world we have that $w \in \mathcal{E}(t, \varphi)$, then term t is *admissible* evidence for φ at world w. Now from definition 2.9 we see that having evidence for φ is not enough for knowledge of it, we need also all our accessible worlds to satisfy φ. We may fail to know something because the reason we might have for believing it is not appropriate, or because it is simply not believable. In order to have *conclusive* evidence for some proposition, both conditions need to be satisfied.[25]

Now the workings of the additional operators on terms will get clear when stating the (possible) axioms of justification logics.

Definition 5.8. *(Axioms and rules of justification logic LP)*[26]

- *A complete axiomatisation of classical propositional logic;*
- *Application axiom:* $\vdash s : (\varphi \to \psi) \to (t : \varphi \to (s \cdot t) : \psi)$*;*
- *Monotonicity axiom:* $\vdash s : \varphi \to (s + t) : \varphi$; $\vdash t : \varphi \to (s + t) : \varphi$*;*
- *Factivity axiom:* $\vdash t : \varphi \to \varphi$*;*

[23] This definition follows Halpern and Pucella 2007.
[24] This definition follows Artemov and Fitting 2016.
[25] Terminology following Artemov and Fitting 2016.
[26] Definition following Artemov and Kuznets 2009.

- *Positive Introspection axiom:* $\vdash t : \varphi \rightarrow !t : (t : \varphi)$;
- *Axiom Internalisation rule: for any logical axiom A and justification constant e, $e : A$*

The application axiom captures the workings of '$\cdot$' to be just like modus ponens: if we have some justification t for some formula φ, and a justification s for φ implying some other formula ψ, then the application of s to t is a justification for ψ. The monotonicity axiom assures that if we have a justification for some formula, and we add some more evidence, it still is a justification of that formula. The positive introspection axiom says that if an agent has a justification for some formula, she can also produce a justification for having a justification of that formula. The axioms of classical propositional logic, the application axiom and the monotonicity axiom hold for all justification logics that we consider. The last three lines of the axiomatisation do not hold in general.

The system LP is the justification logical counterpart of epistemic (modal) logical system **S4**. The relation between them is called the 'forgetful projection', relating $t : \varphi$ to $K\varphi$ by 'forgetting' justification t. Clearly, this means that there may be (and there are) multiple justification logical counterparts of one epistemic logical system. The epistemic logical systems **K**, **T**, **S4** and **S5** that were discussed previously in this chapter, all have justification logical counterparts.[27]

Often justification logics are assumed to satisfy the logical awareness principle, which means that the agent accepts all logical axioms as justified. Of course, this leads to the first form of logical omniscience, closure under theoremhood LO(1). However, we can avoid this by letting our constant specification reflect the agents logical acumen. The constant specification $\mathcal{CS}$ of a justification logic is the set of all statements of the form $e : A$ in which e is a piece of evidence (justification constant) that justifies axiom A of our underlying logic.[28] Formally,

Definition 5.9. *(Constant specification for a justification logic JL)*[29]
$\mathcal{CS} \subseteq \{e : A \mid A \textit{ is an axiom of JL and } e \textit{ a justification constant}\}$

Now let $\mathcal{CS}(e)$ be the set that consists of all the axioms that e justifies. Now we could define our constant specification to be 'axiomatically appropriate[30]': $\bigcup\limits_{e} \mathcal{CS}(e) = \{A \mid A \text{ is an axiom of JL}\}$.

Impossible worlds semantics

The impossible worlds model and knowledge in it are defined as follows:

Definition 5.10. *Let $\mathbf{M} = (W, N, R, \|\cdot\|, \mathcal{C})$ be an epistemic impossible worlds model*[31] *with W the non-empty set of epistemically possible worlds, N the set of 'normal' or logically possible worlds and $\|\cdot\|$ the usual valuation function in standard epistemic logic (for normal worlds) and $\mathcal{C} : W \backslash N \mapsto \mathcal{P}(\mathcal{L})$ listing all known propositions for each (non-normal) world, for $\mathcal{L}$ the language based on the set of atoms At. Valuation is defined with the help of auxiliary functions $\|\cdot\|$ and $\mathcal{C}(w)$, for normal and non-normal worlds respectively:*

- $\forall w \in N$, $M, w \models p$ *iff* $w \in \|p\|_M$;
- $\forall w \in N$, *the truth conditions of the connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$ are the same as in standard epistemic logic;*
- $\forall w \in N$, $M, w \models K\varphi$ *iff* $M, w' \models \varphi$ *for all* $w' \in W$ *such that* wRw';
- $\forall w \in W \backslash N$, $M, w \models \varphi$ *iff* $\varphi \in \mathcal{C}(w)$.

[27] These are named, respectively: J, AJ, AJ4 or LP, and AJ45.
[28] Usually, classical propositional logic is chosen to underlie justification logic, see definition 2.10.
[29] Definition following Artemov and Kuznets 2009.
[30] Definition following Artemov and Kuznets 2009.
[31] This definition follows Halpern and Pucella 2007.

Neighbourhood semantics

Definition 5.11. [32] *Let W be a non-empty set (of possible worlds). A function $N : W \to \mathcal{P}(\mathcal{P}(W))$ is called a* ***neighbourhood function****. A* ***neighbourhood model*** *is a tuple $(W, N, ||\cdot||)$, such that W a non-empty set, N is a neighbourhood function defined over W, and $||\cdot|| : At \to \mathcal{P}(W)$ is a valuation function that assigns a set of possible worlds to each atomic proposition.*

The truth conditions of neighbourhood semantics can be seen to be quite similar to those for normal modal logics:

Definition 5.12. *Let $M = (W, N, ||\cdot||)$ be a neighbourhood model, and $w \in W$. Let $||\varphi||_M$ denote the truth set of φ under the valuation of model M: $||\varphi||_M := \{w \mid M, w \models \varphi\}$. Truth of a formula φ of our language $\mathcal{L}$ is recursively given by the following rules:*

1. $M, w \models p$ *iff* $w \in ||p||_M$
2. $M, w \models \neg\varphi$ *iff* $M, w \not\models \varphi$
3. $M, w \models \varphi \wedge \psi$ *iff* $M, w \models \varphi$ *and* $M, w \models \psi$
4. $M, w \models \Box\varphi$ *iff* $||\varphi||_M \in N(w)$
5. $M, w \models \Diamond\varphi$ *iff* $(W - ||\varphi||_M) \notin N(w)$

For our present purpose we interprete the neighbourhood function epistemically as giving us the known sentences in a world:

$M, w \models K\varphi$ iff $||\varphi||_M \in N(w)$

[32] Definitions following Pacuit 2007.

Appendix C

This appendix shows the questions that served as a starting ground for the interviews conducted for this thesis project.

Interview Questions

1. **Aim of epistemic logic**: What, do you think, are or should be the aims of epistemic logic? Furthermore, is epistemic logic more descriptive, or more normative?
2. **Claims made**: Of what nature are the claims that are made in epistemic logic? Are they generally of a more conceptual nature, or of an empirical nature?
3. **Succes**: When, do you think, is an epistemic logical theory 'successful? What role should meta-logical considerations like soundness, completeness, decidability, et cetera, play, and what role empirical data? How do these different considerations connect, and which type of considerations are most important? Are there perhaps yet other criteria or considerations that can be essential to the success of a theory?
4. **Connection to human knowledge**: What is the object of study of epistemic logic, and how do formal models relate to it? What connection does epistemic logic have to real human agents and our everyday conception of knowledge? Which role do and should intuitions about knowledge play in epistemic logic? Which role are they not allowed to play?
5. **Idealisations**: Do idealisations like logical omniscience constitute a problem for epistemic logic? Why, or why not? If they constitute a problem, how can or should this be resolved?

Bibliography

Achinstein, P. (1964). Models, analogies, and theories. *Philosophy of Science 31*(4), 328–350.

Ackermann, R. J. (1972). *Belief and Knowledge.* NY: Anchor Books.

Artemov, S. and M. Fitting (2016). Justification logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.

Artemov, S. and R. Kuznets (2009). Logical omniscience as a computational complexity problem. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pp. 14–23. ACM.

Artemov, S. and R. Kuznets (2014). Logical omniscience as infeasibility. *Annals of pure and applied logic 165*(1), 6–25.

Aucher, G. (2010). An internal version of epistemic logic. *Studia Logica 94*(1), 1–22.

Baltag, A., L. S. Moss, and S. Solecki (2016). The logic of public announcements, common knowledge, and private suspicions. In *Readings in Formal Epistemology*, pp. 773–812. Springer.

Baltag, A. and S. Smets (2010). Multi-agent belief dynamics. *course given at NASSLLI*.

Baltag, A. and S. Smets (2016). A qualitative theory of dynamic interactive belief revision. In *Readings in Formal Epistemology*, pp. 813–858. Springer.

Berto, F. (2013). Impossible worlds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 ed.). Metaphysics Research Lab, Stanford University.

Blackburn, P., J. F. van Benthem, and F. Wolter (2006). *Handbook of modal logic*, Volume 3. Elsevier.

Cartwright, N. (1983). *How the laws of physics lie.* Oxford University Press.

Cartwright, N. (1992). Aristotelian natures and the modern experimental method. *Inference, Explanation and Other Philosophical Frustrations*, 44–71.

Cartwright, N. et al. (1989). *Nature's Capacities and their Measurement.* Oxford University Press.

Castañeda, H.-N. (1964). Hintikka jaakko. knowledge and belief. an introduction to the logic of the two notions. cornell university press, ithaca, ny, 1962, x+ 179 pp. *The Journal of Symbolic Logic 29*(03), 132–134.

Chisholm, R. M. (1963). The logic of knowing. *The journal of philosophy 60*(25), 773–795.

Chisholm, R. M. (1973). *The problem of the criterion.* Marquette University Press Milwaukee.

Cummins, R. C. (1989). Meaning and mental representation.

Da Costa, N. C. and S. French (2003). *Science and partial truth: A unitary approach to models and scientific reasoning.* Oxford University Press.

De Finetti, B. (1972). Probability, induction, and statistics.

Demey, L., V. F. Hendricks, O. Roy, and D. Pritchard (2011). *Epistemic Logic. 5 Questions.* JSTOR.

Ditmarsch, H., J. Y. Halpern, W. van der Hoek, and B. P. Kooi (2015). *Handbook of epistemic logic.*

Downes, S. M. (1992). The importance of models in theorizing: A deflationary semantic view. In *PSA: Proceedings of the biennial meeting of the philosophy of science association*, Volume 1992, pp. 142–153. Philosophy of Science Association.

Dretske, F. (1981). *Knowledge and the Flow of Information.* MIT Press.

Duhem, P. (1962). The aim and structure of physical theory, trans. by philip p. *Wiener. Athenaeum, New York.*

Eddington, A. (1929). *Scientific Books: The Nature of the Physical World.* New York: Macmillan.

Fagin, R. and J. Y. Halpern (1987). Belief, awareness, and limited reasoning. *Artificial intelligence 34*(1), 39–76.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Vardi (1995). *Reasoning about knowledge.* MIT press.

Frege, G. (1980). Über sinn und bedeutung (on sense and reference). translated by p. geach and m. black.

Frigg, R. (2010). Models and fiction. *Synthese 172*(2), 251–268.

Frigg, R. and S. Hartmann (2017). Models in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.). Metaphysics Research Lab, Stanford University.

Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese 140*(1), 97–119.

Garber, D. (1983). Old evidence and logical omniscience in bayesian confirmation theory. *Testing scientific theories 10*, 99–131.

Garson, J. (2016). Modal logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 ed.). Metaphysics Research Lab, Stanford University.

Georgacarakos, G. et al. (1976). Semantics for $s4.04$, $s4.4$, and $s4.3.2$. *Notre Dame Journal of Formal Logic 17*(2), 297–302.

Gettier, E. L. (1963). Is justified true belief knowledge? *analysis 23*(6), 121–123.

Giere, R. N. (1988). *Explaining science: A cognitive approach.* University of Chicago Press.

Giere, R. N. (1999). *Science without laws.* University of Chicago Press.

Gochet, P. and P. Gribomont (2006). Epistemic logic. *Handbook of the History of Logic 7*, 99–195.

Godfrey-Smith, P. (2009). Models and fictions in science. `http://nrs.harvard.edu/urn-3:HUL.InstRepos:3168793`.

Groenewold, H. (1961). The model in physics. In *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, pp. 98–103. Springer.

Halpern, J. Y. and R. Pucella (2007). Dealing with logical omniscience. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pp. 169–176. ACM.

Hao, W. (1986). *Beyond Analytic Philosophy: Doing Justice to What We Know.* MIT Press.

Hendricks, V. F. (2005). *Mainstream and formal epistemology.* Cambridge University Press.

Hendricks, V. F. and J. Symons (2006). Where's the bridge? epistemology and epistemic logic. *Philosophical Studies 128*(1), 137–167.

Hintikka, J. (1962). *Knowledge and belief: an introduction to the logic of the two notions*, Volume 4. Cornell University Press Ithaca.

Hintikka, J. (1970). 'knowing that one knows' reviewed. *Synthese 21*(2), 141–162.

Hintikka, J. (1979). Impossible possible worlds vindicated. *Game-Theoretical Semantics*, 367–379.

Hintikka, J. (1986). Reasoning about knowledge in philosophy: the paradigm of epistemic logic. In *TARK*, Volume 86, pp. 63–80.

Hintikka, J. (2003). A second generation epistemic logic and its general significance. In *Knowledge contributors*, pp. 33–55. Springer.

Hocutt, M. O. et al. (1972). Is epistemic logic possible? *Notre Dame Journal of Formal Logic 13*(4), 433–453.

Huang, Z. and K. Kwast (1990). Awareness, negation and logical omniscience. In *European Workshop on Logics in Artificial Intelligence*, pp. 282–300. Springer.

Hughes, R. I. (1997). Models and representation. *Philosophy of science 64*, S325–S336.

Jones, M. R. (2005). Idealization and abstraction: A framework. *Idealization XII: Correcting the model. Idealization and abstraction in the sciences 86*, 173–217.

Kasbergen, O. (2012). On the surprise examination. `https://esc.fnwi.uva.nl/thesis/centraal/files/f1882374659.pdf`.

Klein, D. (2015). *Social Interaction: a Formal Exploration.* Ph. D. thesis, uitgever niet vastgesteld.

Kripke, S. (1980). Naming and necessity harvard university press cambridge. *MA: Harvard University Press*.

Kurucz, A., F. Wolter, M. Zakharyaschev, and D. M. Gabbay (2003). *Many-dimensional modal logics: theory and applications*, Volume 148. Gulf Professional Publishing.

Lehman, R. S. (1955). On confirmation and rational betting. *The Journal of Symbolic Logic 20*(03), 251–262.

Lemmon, E. (1967). Mathematics and logic in history and contemporary thought.

Lenzen, W. (1978). Recent work in epistemic logic.

Lenzen, W. (2004). Epistemic logic. In *Handbook of Epistemology*, pp. 963–983. Springer.

Lloyd, E. A. (1994). *The structure and confirmation of evolutionary theory.* Princeton University Press.

Mattey, G. J. (2015, April). Introduction to theory of knowledge. http://hume.ucdavis.edu/mattey/phi102/intro.html#Projects.

McMullin, E. (1984). *A case for scientific realism.*

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A 16*(3), 247–273.

Menzel, C. (2016). Possible worlds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.

Morgan, M. S. (2001). Models, stories and the economic world. *Journal of Economic Methodology 8*(3), 361–384.

Morrison, M. and M. S. Morgan (1999). Models as mediating instruments. *Ideas in context 52*, 10–37.

Nozick, R. (1981). *Philosophical Explanations.* Harvard University Press.

Pacuit, E. (2007). Neighborhood semantics for modal logic: An introduction. *Course Notes for ESSLLI 2007.*

Parikh, R. (1987). Knowledge and the problem of logical omniscience. In *ISMIS*, Volume 87, pp. 432–439.

Rantala, V. (1979). Urn models: a new kind of non-standard model for first-order logic. *Game-Theoretical Semantics*, 347–366.

Rescher, N. (2005). *Epistemic logic: a survey of the logic of knowledge.* University of Pittsburgh Pre.

Rescher, N. (2006). Epistemic logic. In D. Jacquette (Ed.), *A companion to philosophical logic*, Chapter 30, pp. 478–490. Blackwell Publishing Ltd.

Scott, D. (1973). Background to formalization. *Studies in Logic and the Foundations of Mathematics 68*, 244–273.

Stalnaker, R. (1991). The problem of logical omniscience, i. *Synthese 89*(3), 425–440.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical studies 128*(1), 169–199.

Stokhof, M. (2012). The role of artificial languages. *The Routledge companion to the philosophy of language. Routledge, London/New York*, 544–553.

Stokhof, M. and M. Van Lambalgen (2011). Abstractions and idealisations: The construction of modern linguistics. *Theoretical linguistics 37*(1-2), 1–26.

Suárez, M. (1999). Theories, models, and representations. In *Model-based reasoning in scientific discovery*, pp. 75–83. Springer.

Suppes, P. (1961). A comparison of the meaning and uses of models in mathematics and the empirical sciences. In *The concept and the role of the model in mathematics and natural and social sciences*, pp. 163–177. Springer.

Tarski, A. (1953). *Undecidable theories*, Volume 13. Elsevier.

Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis 55*(3), 393–415.

Van Benthem, J. (2006). Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies 128*(1), 49–76.

Van Benthem, J. (2007). Logic in philosophy. In D. Jacquette (Ed.), *Handbook of the Philosophy of Science*, Chapter 4, pp. 65–99. Elsevier.

Van Benthem, J. (2010). *Modal logic for open minds.* Center for the Study of Language and Information.

Van der Hoek, W. (1993). Systems for knowledge and belief. *Journal of Logic and Computation 3*(2), 173–195.

Van Ditmarsch, H., J. Y. Halpern, W. Van Der Hoek, and B. Kooi (2015). An introduction to logics of knowledge and belief. *arXiv preprint arXiv:1503.00806*.

Van Fraassen, B. C. (1980). *The scientific image.* Oxford University Press.

Von Kutschera, F. (1976). Einführung in die intensionale semantik.

Weisberg, M. (2007a). Who is a modeler? - final draft. `http://www.phil.upenn.edu/~weisberg/documents/modelingfinal.pdf`.

Weisberg, M. (2007b). Three kinds of idealization. *The journal of Philosophy 104*(12), 639–659.

Williamson, T. (2002). *Knowledge and its Limits.* Oxford University Press on Demand.

Williamson, T. (2013). Knowledge first. In M. S. J. Turri (Ed.), *Contemporary Debates in Epistemology*, pp. 1–10. Blackwell.

Wiśniewski, A. (1998). Two logics of occurrent belief. *Acta Universitatis Wratislaviensis 2023.*