Utrecht University

Bachelor Artificial Intelligence

Desired results for an epistemic logic for bounded agents

Author
Sanderijn Kuijvenhoven
6227295

Supervisor
Colin Caret
Second reader
Natasha Alechina

Bachelor Thesis 7,5 ECTS
6 November 2020

Abstract

The aim of this thesis is to suggest beneficial results when constructing an epistemic logic for bounded agents. To model human knowledge in machines, knowledge must first have a correct representation. Epistemic logic tries to represent knowledge but the standard system, the S5 system, comes with many problems. The biggest problem is that the system requires its agents to be logically omniscient [**van Ditmarsch et al.**, **2007**]. To avoid problems like these, this thesis suggests properties which are desirable to implement for an epistemic logic for bounded agents. The properties this thesis discusses are factivity, order of conjunction, and the difference between agents in reasoning capacity. With these properties two different formal approaches to epistemic logic are analysed, the awareness logic [**Fagin et al.**, **1987**] and the impossible possible worlds semantics [**Hintikka**, **1979**]. However, both of these approaches satisfied the same properties; both did not necessarily satisfy factivity, both did not necessarily satisfy the order of conjunction, and both did satisfy the difference between agents. The desirable results an epistemic logic would want to have are factivity, the order of conjunction, and a way to show the difference between agents in reasoning capacity. Awareness logic and impossible possible worlds semantics need some adaptations in order to have beneficial results. Further research could be done to find more beneficial properties for an epistemic logic to implement, to see how other approaches to epistemic logic perform using the properties, and to see how humans think about knowledge.

Keywords: logical omniscience, bounded agents, awareness logic, impossible possible worlds semantics

Contents

1 Introduction

Epistemology is the study of knowledge. It is useful to model human knowledge into an epistemic logic. Logic is the study of reasoning, whereby epistemic logic approaches to reasoning about knowledge, belief, and related notions [**Rendsvig et al.**, **2019**]. The knowledge of human or artificial agents can be represented using formal mathematical tools. Representing knowledge in logic can be useful for reasoning about knowledge and for reasoning about other agents with knowledge.

Humans do not have an infinite reasoning capability and can therefore be seen as bounded agents [**van Linder et al.**, **August 1998**]. Within the standard knowledge representation in logic, some problems occur when bounded agents are analysed instead of agents with an infinite amount of reasoning capabilities. The standard system for knowledge, the S5 system, is closed under logical consequence [**Sim**, **1997**]. This means that the system assumes that the agents can know all logical consequences and thus assumes that agents are unbounded to reasoning capacity or capability. This problem is also known as logical omniscience [**van Ditmarsch et al.**, **2007**], since the system requires agents to be logical omniscient. This thesis will look at properties an epistemic logic for bounded agents would benefit from. The research question this paper will try to answer is:

What are the desired results for constructing an epistemic logic for bounded agents?

To try to answer this question each chapter will look at a different aspect of the question and answer a sub-question related to the research question. The sub-questions this thesis will answer are:

1. What is the standard approach to represent knowledge in epistemic logic?
2. What problems occur within this standard approach to epistemic logic?
3. What are concrete (philosophical) examples for properties that would be beneficial to implement into an epistemic logic?
4. Which formal epistemic system represents knowledge for bounded agents more beneficial, when keeping the examples from chapter 3 in mind?

This thesis will analyse the standard approach to represent knowledge in epistemic logic and the problems that arise from using this approach. The main focus of this thesis is to discuss what properties would be beneficial to an epistemic logic. It will discuss some concrete examples of what would be desired properties for an epistemic logic. With these concrete examples the paper will then compare two formal solutions, with different approaches, to solve the problems of the standard approach to epistemic logic. Important literature this thesis will look into is the book *Dynamic epistemic logic* [**van Ditmarsch et al.**, **2007**], the book *reasoning about knowledge* [**Fagin et al.**, **2003**], and the article *impossible worlds vindicated* [**Hintikka**, **1979**].

1.1 Relevance to Artificial Intelligence

Before trying to answer the research question, the scientific relevance of this thesis must be addressed, specifically the relevance to Artificial Intelligence. Artificial Intelligence is an area within computer science which tries to mimic human-reasoning and human-actions by using computers [**Meyer et al.**, **2004**]. By recreating human intelligence, it

is important to know how to represent human-knowledge in a correct way if it has to be implemented in computers. If human-knowledge can not be correctly represented, it is impossible to implement knowledge to computers. Besides this, it is also important for humans to know how we reason so we can reason about other agents, this can be humans as well as computers. But besides humans reasoning about other agents, computers need to reason about humans as well. When an Artificial Intelligence needs to reason about human knowledge, it is important that the AI does not make any mistakes because the knowledge is wrongly implemented[**Sim**, **1997**].

Next to having to implement human reasoning and knowledge to machines, this paper also looks at bounded agents. In the previous sections the most obvious example of bounded agents is introduced: humans. But humans are not the only agents that can be seen as bounded agents. Van Linder et al. [**1998**] describe Artificial agents as soft-bots and robots. Machines and computers fall into the artificial agents category, since they have a finite amount of memory space and thus are limited to this space. So, computers and machines can also be seen as bounded agents because they do not have infinite memory. When computers are used as bounded agents it is important that when applying epistemic logic, no problems occur. For example, it must be avoided that computers have a knowledge base that assumes that agents are logically omniscient.

1.2 Structure of the paper

To answer the research question: "What are the desired results for constructing an epistemic logic for bounded agents?". This paper will contain three chapters with each different aspect related to the research question. Each chapter will answer a sub-question to help answer the main research question. Firstly, in chapter 2 the basic approach to representing knowledge in epistemic logic will be analysed and explained. The sub-question, "what is the standard approach to represent knowledge in epistemic logic?", will be answered by introducing what the basic approach to epistemic logic is, and explaining how this system works and how this system is represented. In the second section, the sub-question, "what problems occur within this epistemic logic?", will be answered by analysing, and explaining all the problems that occur within epistemic logic, especially when looking at epistemic logic for bounded agents. In this chapter important literature will be *Epistemic logic and logical omniscience: a survey* [**Sim**, **1997**] and *Dynamic epistemic logic* [**van Ditmarsch et al.**, **2007**].

Secondly, chapter 3 will answer the sub-question "What are concrete (philosophical) examples for properties that would be beneficial to implement into an epistemic logic?". In this chapter different properties will be discussed to see how they could be useful for an epistemic logic and why it would be beneficial to implement them. In this chapter important literature will be *Mythology of the factive* [**Turri**, **2001**] and again *Dynamic epistemic logic* [**van Ditmarsch et al.**, **2007**].

Lastly, chapter 4 will try to answer the sub-question, "which formal epistemic system represents knowledge for bounded agents more beneficial, when keeping the examples from chapter 3 in mind?". Two different approaches to epistemic logic will be explained and analysed using the examples from chapter 3. From analysing these approaches it can be seen if one approach is more beneficial than the other approach. The first approach will be the awareness logic and the second approach will be the impossible possible world semantics. In this chapter important literature will be *Impossible Possible Worlds Vindicated*[**Hintikka**, **1979**], *Belief, awareness, and limited reasoning* [**Fagin et alm 1987**] and *Belief, Awareness, and limited reasoning* [**Fagin et al.**, **2003**]

2 Chapter 2

In this chapter the questions: "what is the standard approach to represent knowledge in epistemic logic?" and "what problems occur within this standard approach epistemic logic? will be answered. The first question will be answered in the first subsection and the second question will be answered in the second subsection.

2.1 Epistemic Logic

Epistemic logic is a sub-field of epistemology and concerns itself with the formal representation of knowledge and belief.

Hintikka's book *Knowledge and Belief: An Introduction to the logic of the Two Notion* [**1962**] was the first book where a clear semantics for knowledge was introduced. He applied possible world semantics to model knowledge and added the notion of accessibility between worlds. The idea of possible world semantics is to think of the information an agent has in terms of possible worlds that are in agreement with the information of the agent. Between these worlds are accessibility relations, this means that an agent knows something, if and only if, it is the case for all possible worlds accessible to the agent. Thus, an agent can not know something if it is not true for all of the worlds accessible to the agent. In other words, $K\varphi$ is true in world w, if and only if, φ is true in every world w' compatible with what an agent knows at w. This is also called the partition principle; any propositional attitude partitions the set of possible worlds into those that are in accordance with the attitude of those that are not [**Rendsvig et al.**, **2019**]. The partition principle may be used to provide a semantics for the knowledge operator. With the possible world semantics it is easy represent knowledge and keep an overview at the same time. Besides that, it is also easy to add more agents to the possible world semantics and show what they know about each other.[**Van Ditmarsch et al.**, **2007**]

2.1.1 Syntax S5 system

Epistemic logic is influenced by Kripke semantics and modal logic, which will both be explained later [**van Ditmarsch et al.**, **2007**]. The most used system in Epistemic logic is S5 and could also be called the "basic" system [**van Ditmarsch et al.**, **2007**]. The smallest normal modal logic is K and the S5 system is the smallest normal modal logic that includes all of the S5 principles, this means that the system contains exactly what is required and nothing more [**Rendsvig et al.**, **2019**]. In epistemic logic, knowledge is represented with the modal operator K. This makes the set of well-formed formulas in S5 as follows: let P be a set of atomic propositions,

$$\varphi := p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K\varphi$$

for $p \in P$. These well-formed formulas can be seen as the building blocks for other formulas because complex formulas can be generated from these building blocks. The modal operator K is also introduced. For every agent a, $K_a\varphi$ is interpreted as *agent a knows that* φ. In this thesis only single agents will be discussed, while the S5 system can also be used for a group of agents [**van Ditmarsch et al.**, **2007**].

2.1.2 Semantics S5 system

At first epistemic logic was represented with propositional logic, but with this approach it was difficult to model all scenario's. For example, take the sentence "it is raining" and imagine that you are in a situation where you do not know and could not tell if it is raining or not. In this situation either one can be true; it is raining or it is not raining. To model this, Hintikka thought of using Kripke semantics and possible worlds semantics to model epistemic knowledge [**Hintikka**, **1962**]. The possible words semantics consists of Kripke structures. These can be divided in Kripke frames and models:

Kripke frame

$$F = \langle W, R \rangle$$

A Kripke frame is a tuple with two elements. Firstly, W is a non-empty set of possible worlds. And secondly, $R \subseteq (W \times W)$ is an equivalence relation on W, this will be further explained in the next section. For example, if $w_1 R w_2$ then that means that w_1 can access w_2. The Kripke frame provides a structure on which a Kripke model can be built.

Kripke model

Given a set of atomic propositions P, a Kripke model is a structure:

$$M = \langle W, R, V \rangle$$

A Kripke model consists of a Kripke frame, $\langle W, R \rangle$, to use as the basis for the Kripke model and V, the valuation function. The valuation function shows which proposition is true in which world. It states that for every $p \in P$ it holds that $V(p) \subseteq W$ of worlds in which p is true [**van Ditmarsch et al.**, **2007**]. The previous section stated that $K\varphi$ is true in world w, if and only if, φ is true in every world w' compatible with the information an agent has at w. The formal symbol for stating that φ is true in world w is $w \models \varphi$ and is often called the satisfaction relation [**Rendsvig et al.**, **2019**]. When representing a Kripke frame, worlds are usually represented with nodes or the name of the world and the accessibility relations are represented using arrows. When a Kripke model is represented, at each world the set of propositions that are true in that world are written next to the node or name representing that world.

To verify whether a formula is true at a given world or not, the modal operator K must be further explained. The operator K can be seen as a $\Box$ operator in modal logic [**Rendsvig et al.**, **2019**]. This means for $w_1 \models K\varphi$ to be true, for every world $w \in W$ such that the access relation $w_1 R w$ holds $w \models \varphi$. There are cases when $\neg K \neg \varphi$ holds, then this makes an existential quantification. This means that there exists an accessible world that satisfies φ. $\neg K \neg \varphi$ is also represented as $\widehat{K}\varphi$ or $\langle K \rangle$ and it acts like the $\Diamond$ operator in modal logic [**Rendsvig et al.**, **2019**]. For $w_1 \models \widehat{K}\varphi$ to be true, it must hold that the accessibility relation must hold from w_1 to some world w_2 where $w_2 \models \varphi$.

The example introduced at the beginning of this paragraph can now be modelled into a Kripke model M. There will be two worlds; w_1 where it is raining and w_2 where it is not raining. The atom p stands for it is raining and $p \in P$ for P is the set of atomic propositions. With truth valuations $M, w_1 \models p$ and $M, w_2 \models \neg p$. This would model as:

To see if $w_1 \models Kp$ holds in this model, all worlds that are accessible from w_1 must have $\models p$. There are two accessibility relations from w_1, w_1 and w_2. Firstly, $w_1 \models p$ so

the first one holds. Secondly, $w_2 \nvDash p$ so the second world does not match the requirement. So it can be concluded that $w_1 \nvDash Kp$.

2.1.3 Proof-theory S5 systems

The S5 system comes with a few epistemological principles or principle which set up the system. Some of these epistemic principles are relations between worlds that must hold. These relations can be shown using the possible worlds semantics. The principles are used as a basis to help verify when a formula is true, mostly in proofs [**van Ditmarsch et al.**, **2007**]. Besides using the principles of S5, the only two other rules of inferences that may be used to prove if a formula is true, are modes ponens and necessitation [**Rendsvig et al.**, **2019**]. Modus ponens is the rule that from φ and $\varphi \to \psi$, follows ψ. Necessitation is the rule that form φ follows $K\varphi$. To make a correct proof, start from a set of assumptions X. With these set of assumptions the epistemological principles can be applied and the two other inference rules can be applied to arrive at the correct ending.

The first epistemological principle is:

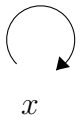
$$\mathrm{K} : K(\varphi \to \psi) \to (K\varphi \to K\psi)$$

K stands for the closure under implication [**van Ditmarsch et al.**, **2007**]. A closure principle means that a set of objects is closed relative to a function or rule [**Kvanvig**, **2006**]. In other words, the function on a member of the set always leads to something already in the set. This closure under implication principle could also be called logical consequence [**Holliday**, **2015**]. It states that if an agent knows that if φ is true then ψ must be true, then if that agent knows that φ is true then that agent knows that ψ.

The second principle is:

$$\mathrm{T} : K\varphi \to \varphi$$

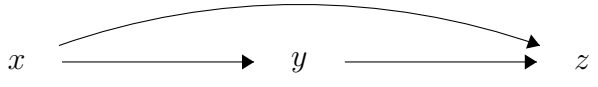
T shows that knowledge is factive. T is also known as the reflexive principle [**Rendsvig et al.**, **2019**]. This principle states that if an agent knows a formula φ then that formula φ is true. This rule makes it possible for an agent to only know a formula when that formula is true. Thus, an agent cannot know a formula if that formula is false. Besides, $Kp \to p$, other formulas can also be factive. For example, $K(p \wedge Kq) \to (p \wedge Kq)$. The reflexive principle hold in possible world semantics if $\forall x$ states that xRx, in other words the world has an accessibility relation to itself. The reflexive principle applied with possible world semantics would look like:



The third principle is:

$$4 : K\varphi \to KK\varphi$$

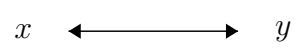
4 shows that knowledge is positive introspective [**van Ditmarsch et al.**, **2007**]. It states that if an agent knows that φ then the agent knows that he or she knows that φ. This means that agents know what they know. The transitivity principle holds in possible world semantics if $\forall x, y, z$ if xRy and yRz, then xRz. This principle applied to possible worlds semantics looks like:



The last epistemic principle is:

$$\mathrm{B} : \varphi \to K\widehat{K}\varphi$$

B shows that the relations must be symmetric within the S5 system [**Rendsvig et al.**, **2019**]. This means that $\forall x, y$ it holds that if xRy then yRx. In other words, if there is an arrow from world x to world y then there must also be an arrow from world y to world x. The symmetric principle can also be applied with possible world semantics and would like like:



A relation that is both reflexive, symmetric and transitive is called an equivalence relation [**Rendsvig et al.**, **2019**]. This means that if an epistemic logic follows the principles of the S5 system, the accessibility relation that holds for K to capture knowledge must be an equivalence relation.

The example from the previous section, is also represented within the S5 system. So all the epistemic principles K, T, and 4 hold in this scenario. w_1 is reflexive because w_1Rw_1 and w_2 is also reflexive because w_2Rw_2. In addition this scenario is also transitive because there is no third world and all the possible accessibility relations already exist in this model.

2.2 Problems S5 system

The S5 system is a good starting place for constructing an epistemic logic for bounded agents. However, the S5 system does come with its problems. Firstly, the system requires an agent to be logical omniscient. Logical omniscience is the notion of logical implication [**Fagin et al.**, **2003**]. The notion can be viewed as a certain closure property of an agent's knowledge. It says that if an agent knows certain formulas and certain conditions hold then the agent must also know some other formulas [**Fagin et al.**, **2003**]. The requirement of logical omniscience follows from the K principle, $K(\varphi \to \psi) \to (K\varphi \to K\psi)$. This principle shows that knowledge is closed under implication. Because of this closure property a formula such as:

$$K\varphi \wedge K(\varphi \to \psi) \to K\psi$$

is true when applying the K principle. This formula states that if an agent knows that φ is true and ψ follows logically from φ, then the agent must therefore also know ψ. However, the principles closure property is too strong when human reasoning capabilities are being analysed. It can not be assumed that humans know the logical consequence of a logical truth because humans have a limited reasoning capacity. A great example for this is mentioned in the book *Reasoning about knowledge* [**Fagin et al.**, **2003**], a person can know the rules of chess without knowing whether or not White has a winning strategy. Therefore it may not be assumed that humans can know all logical consequences, since humans are bounded agents. The S5 system requires more then the human reasoning capabilities [**van Ditmarsch et al.**, **2007**]. Still there is a deviation in reasoning capabilities between humans, this will be discussed in the next chapter. The logical omniscience problem show that the S5 system is not the most suitable system for representing knowledge and belief. The logical omniscience problem is one of the most researched problems within epistemic logic [**Holliday**, **2005**]. There have been many attempts at

constructing solutions for this problem. In the fourth chapter of this thesis, two examples of possible solutions will be discussed. While there are many more solutions to this problem, this thesis does not have the space to analyse more solutions.

Besides the closure of implication, the S5 system has more problems. Namely, the following statements are all assumed true within the S5 system [**van Ditmarsch et al.**, **2007**]. Let φ, ψ be formulas in L and let K be an epistemic operator. Let $\mathcal{K}$ be the set of all Kripke models and S5 the set of Kripke models in which the accessibility relation is an equivalence relation. Firstly,

$$\mathcal{K} \models \varphi \Rightarrow \models K\varphi$$

this statement claims that the agent knows all tautologies, it says that if $\models \varphi$ then $\models K\varphi$. This statement follows from the assumption that an agent know all consequences of their beliefs [**Fagin et al.**, **2003**]. However, this is impossible when the agents are bounded agents, such as humans. Agents with limited memory and reasoning capacity can, because of their limitations, not know all tautologies in propositional logic.

$$\mathcal{K} \models \varphi \rightarrow \psi \Rightarrow \models K_a\varphi \rightarrow K_a\psi$$
$$\mathcal{K} \models \varphi \leftrightarrow \psi \Rightarrow \models K_a\varphi \leftrightarrow K_a\psi$$
$$\mathcal{K} \models (K_a\varphi \wedge K_a\psi) \rightarrow K_a(\varphi \wedge \psi)$$
$$\mathcal{K} \models K_a\varphi \rightarrow K_a(\varphi \vee \psi)$$

The statements above are all flaws within the S5 system to represent knowledge [**van Ditmarsch et al.**, **2007**]. All of these statements assume that the agent is able to make logical deductions with respect to his knowledge. The agent is assumed to know all logical implications and connections.

3 Chapter 3

It has been shown that the S5 system has some serious problems. This shows that the principles of the S5 system are too strong for knowledge representation, since it is not representative for bounded agents. But when a system does not have any restrictions or principles that should be satisfied to represent knowledge, knowledge becomes an arbitrary idea where there can not be any logical rules. The question comes up, how much restrictions or principles should be applied to represent knowledge. And where on the spectrum between the arbitrary knowledge and the too strong knowledge representation, should the knowledge represented system be? [**Jago**, **2006**] This chapter argues for a three simple principles that the representation of knowledge would want to implement to be in between the two ends and tries to answer the sub-question "What are concrete (philosophical) examples for properties that might be beneficial to implement into an epistemic logic for bounded agents?". This question will be answered by suggesting different examples for desirable properties that an epistemic logic for bounded agents might want to have.

3.1 Factivity

The most basic property an epistemic logic might want to implement is factivity of knowledge. The principle of factivity is that if an agent knows a proposition, that proposition is true. In other words, knowledge implies truth.

$$K\varphi \rightarrow \varphi$$

In the previous chapter this principle was also introduced as the reflexive principle T. The factivity principle states that an agent can only know a proposition when that proposition is true [**Ichikawa et al.**, **2018**]. So this means that every statements known by an agent is true, since agents can only know true statements. If an agent thinks he or she knows a proposition and this proposition is false then the agent can not know this proposition.

This is principle is a very common idea and most epistemologists have found it overwhelmingly plausible that what is false cannot be known [**Ichikawa et al.**, **2018**]. In his article *Mythology of the factive* John Turri [**2011**] claims that the position of factivity in knowledge can adequately and uniformly explain all cases of why non-factive uses of knowing sound odd. A good example Turri mentions of a case where non-factive knowing sounds odd is: Dick is asked what he knows about Iran's nuclear program. Condi responds "Dick knows that Iran has built an nuclear bomb, although they have not built one". The response of Condi sounds odd because she contradicts herself. Condi says Dick knows, but in the same sentence denies that he does, because the necessary condition of Dick's knowing is not satisfied. Most of the proof for this principle are examples of situation whereby it feels intuitively wrong for knowledge not to be factive. [**Turri**, **2011**]

When adapting knowledge to computers, it is convenient to state that what humans know is true. It is difficult for a machine to decide for themselves whether or not what a human says is true. This is made easier when knowledge implies truth.

3.2 Order of conjuction

An other basic property an epistemic logic might want to implement in its system, is the order of conjunction. The order of conjunction means that the order of conjuncts within a conjunction should not matter. When looking at an example: $K(p \wedge q) \rightarrow K(q \wedge p)$ and

$K(q \wedge p) \rightarrow K(p \wedge q)$, in both examples it seems clear that the one implies the other. This means that the order of the conjuncts in the conjuction does not matter to the definition or to the truth. SInce $p \wedge q$ and $q \wedge p$ should mean the same thing in the epistemic logic, it can be stated that $p \wedge q$ is equivalent to $q \wedge p$. It seems clear that, for example if an agent knows that there is a cat and a dog, then the agent knows that there is a dog and that there is a cat. The principle for the order of conjunction the epistemic system might want to implement is

$$K(p \wedge q) \rightarrow K(p \wedge q)$$

3.3 Difference between agents

As already mentioned in the previous chapter, an aspect which needs more thought, is the question if knowledge representation should be the same for all aspects. Some agents have better reasoning capabilities than other agents. It is beneficial to implement the difference between agents in an approach to epistemic logic because it makes it easier for agents to reason about other agents knowledge.

To say more about this, there first need to be a definition of an agent. Throughout this thesis the term agent has often been used, but what exactly is an agent? In the article *Formalising abilities and opportunities of agents* it is said that there is no agreement on what the term 'agents' exactly means.[**van Linder et al.**, **1998**]. Later in their paper they form a definition of an agent themselves: "an entity which has the possibility to execute certain actions and is in the possession of certain information, which allows it to reason about its own and other agents' actions" [**van Linder et al.**, **1998**]. This definition will also be assumed in this thesis. Since there are different kinds of agents, with the most obvious examples humans and computers, it might be beneficial to implement these differences into an epistemic logic. Since every agent has a different function, its reasoning capabilities might depend on that function.

For example, a professor in logic is likely to have more reasoning capacity then his students. The study of Hogan et al. indeed found that students and nonscientists differ from technicians and scientists in drawing conclusions. With the major difference in the emphasis on criteria of empirical consistency or plausibility [**Hogan et al.**, **2001**]. It would be a beneficial aspect to an epistemic logic to make a distinction between reasoning capacities of agents.

Naturally, it would be the easiest to assume that every agent has the same reasoning capabilities. However, this is not the case, an example to see difference between different types of agent is by introspection. There are two kinds of introspection; positive and negative introspection. Firstly, positive introspection:

$$K\varphi \rightarrow KK\varphi$$

This principle was above defined as the epistemic principle 4 for the S5 system. With the positive introspection an agent knows what he knows. Secondly, negative introspection:

$$\neg K\varphi \rightarrow K\neg K\varphi$$

This principle states that the agent knows what the agent does not know. The problem arises when looking at different types of agents. For humans positive introspection seems reasonable. However, negative introspection seems like an unrealistic strong assumption. But when looking at artificial agents both positive and negative introspection make sense [**Van Ditmarsch et al.**, **2007**]. Van Linder et al. state that humans are far

more complex than these artificial agents and humans need more complicated description [**van Linder et al.**, **1998**]. Thus, it makes sense to at least make a difference between different types of agents within an epistemic logic.

However, how to implement a difference in reasoning capabilities between agents is hard. Because how can one measure an agents reasoning capabilities or skills? The reasoning capabilities of an agent are dependent on many different variables. And how is decided which variables do have an impact or on an agents reasoning capacity? This is a difficult question for humans, as already said above humans are far more complex than artificial agents. However, for these artificial agents it is easier to decide whether there is a difference between reasoning capacities [**van Linder**]. For example, a computer with a memory of 16 GB and a computer with a memory of only 4 GB. The computer with 16 GB is less limited in his capabilities than the computer with 4GB of memory. An easy approach for how to implement a principle for the difference between agents, is to make a variable for only one aspect that decides the reasoning capabilities between agents. For example a time variable, which is a part of dynamic epistemic logic. Alechina et al. used this method when they constructed an logic for recourse-bounded agents [**Alechina et al**, **. July 2002**]. They made a variable for a computational delay, this delay represents the time it takes for an agent to arrive at the conclusion of logical consequence. This could be an approach to how to implement a principle for the difference between agents, but it is only focused on the time it takes to get to an conclusion. However, an agents reasoning capability is not decided by just one variable.

4 Chapter 4

The possible worlds model is a very useful tool to represent knowledge, but as mentioned in chapter 3, it does come with some serious problems. This chapter will try to answer the sub-question "which formal epistemic system represents knowledge for bounded agents more beneficial, when keeping the examples from chapter 3 in mind?". To answer this question, two different approaches to change the S5 system will be introduced: awareness logic and impossible possible worlds semantics. Firstly, the two approaches will be explained. Then the approaches will be checked with the examples from chapter 3 to see how well the approaches satisfy them. Lastly, the two approaches will be compared to see which of the two approaches represents knowledge more beneficial. Besides the awareness logic and the impossible possible worlds, much more solutions have been thought of by epistemologists. However, there is not enough space here to look at other solutions besides these two.

4.1 Awareness logic

The first approach to epistemic logic this thesis will analyse is the logic of awareness. The idea of awareness logic is that it is necessary to be aware of a concept before the concept can be known by an agent [**Fagin et al.**, **1987**]. Thus, an agent can not know something if he or she is unaware of it. With awareness logic a new model operator is introduced, A_i for each agent i. $A_i\varphi$ is read as i is aware of φ. What makes awareness logic attractive is its flexibility [**Fagin et al.**, **2003**], which will be further explained below.

Awareness logic can be seen as an extension of Levesque's logic. Levesque's logic consist of different kinds of knowledge; K_i for implicit knowledge of agent i and X_i for explicit knowledge of agent i [**Levesque**, **August 1984**]. Firstly, implicit knowledge are all formulas an agent in a position to know, if they had unbounded logical capabilities. In the possible worlds semantics it is this the truth in all worlds that the agent considers as possible. Secondly, explicit knowledge is a combination of implicit knowledge and an extra requirement. Levesque states that this requirement is that the formula is actively held true by the agent [**Levesque**, **August 1984**]. The awareness logic extends Levesque's logic by adding the extra awareness structure to create an awareness model $\mathcal{N} = \langle W, R, A, V\rangle$. The requirement for explicit knowledge in awareness logic is the agent must implicitly know the formula and the agent must be aware of the formula [**Fagin et al.**, **2003**].

An awareness structure is a tuple $M = (S, \pi, \mathcal{K}_1, ..., \mathcal{K}_n, \mathcal{A}_1, ..., \mathcal{A}_n)$, where the tuple $(S, \pi, \mathcal{K}_1, ..., \mathcal{K}_n,)$ is a Kripke structure and $\mathcal{A}(s)$ is a function for the formulas of which agent i is aware. The formulas in $\mathcal{A}(s)$ are formulas which the agent is aware of but not necessarily knows. These formulas can be arbitrary and inconsistent. For example, it is possible for both φ and $\neg\varphi$ to be in $\mathcal{A}(s)$, or might be that $\varphi \wedge \psi$ is in $\mathcal{A}(s)$ but $\psi \wedge \varphi$ is not in $\mathcal{A}(s)$.

With the new modal logic operators in awareness logic, new clauses need to be added for formulas of the form $A_i\varphi$ and $X_i\varphi$ [**Fagin et al.**, **2003**]. Firstly,

$$(M, s) \models A_i\varphi \text{ iff } \varphi \in \mathcal{A}_i(s)$$

This states that agent i is aware of φ at state s, if and only if, φ is in $\mathcal{A}_i(s)$. Secondly,

$$(M, s) \models X_i\varphi \text{ iff } (M, s) \models A_i\varphi \text{ and } (M, s) \models K_i\varphi$$

This states that agent i explicitly knows φ, if and only if, agent i is aware of φ and agent i implicitly knows φ. This shows again that an agent can not explicitly know a formula if the agent is not aware of that formula. From this follows that $X_i \Leftrightarrow A_i\varphi \wedge K_i\varphi$. The implicit operator K behaves the same way as in a Kripke structure. However, the explicit operator X behaves differently. There are for now no restrictions on the formulas of which agents can be aware of [**Fagin et al.**, **2003**]. This means that agents do not have to be aware of logically valid formulas. So, logical omniscience is not a problem in awareness logic, since awareness logic is not closed under logical consequence and there are no further restrictions on the awareness logic. This allows that explicit knowledge does not require its agents to be logical omniscient.

4.1.1 Examples applied

The examples for properties of the previous chapter are used to see how well the awareness logic represents knowledge. The first property is factivity. For awareness logic to be factive, the explicit knowledge has to be factive. This means when an agent explicitly knows a formula then that formula must be true. The example *Reasoning about knowledge* mentions is, assume that $\mathcal{K}_i$ is reflexive so that $K_i\varphi \Rightarrow \varphi$ is valid and $X_i\varphi \Rightarrow K_i\varphi$ is valid, then we obtain the principle $X_i\varphi \Rightarrow \varphi$ [**Fagin et al.**, **2003**]. So, for awareness logic to be factive, the reflexive principle must hold. This means that factivity is not necessarily satisfied in awareness logic since a condition must be hold before factivity can hold in awareness logic.

The second property is the order of conjuction. The order of conjunction is not satisfied in awareness logic. As stated above, the formulas within the awareness function can be arbitrary and consistent [**Fagin et al.**, **2003**]. An agent can not explicitly know a formula when the agent is not aware of that formula. Because of this, it is possible that an agent is aware of $\varphi \wedge \psi$ and implicitly knows $\varphi \wedge \psi$ and therefore explicitly knows $\varphi \wedge \psi$. However, the agent is could not be aware of $\psi \wedge \varphi$ and therefore the agent can not explicitly know $\psi \wedge \varphi$. This example shows that $\varphi \wedge \psi$ can be in $\mathcal{A}(s)$ but $\psi \wedge \varphi$ does not have to be in $\mathcal{A}(s)$, and thus it clearly shows that the awareness logic does not satisfies the order of conjunction principle.

However, this could be solved with introducing a new class. For this solution the notation for the class 'epistemically equivalent formulas' must be introduced. If P is a formula, let $[P] = \{Q \mid P \text{ and Q are equivalent}\}$. Then they must follow the three rules for equivalence: reflexive $p \in [P]$, symmetric $Q \in [P] \rightarrow P \in [Q]$, and transitive $R \in [Q] \wedge Q \in [P] \rightarrow R \in [P]$. When this class is used in awareness logic, it leads to a constraint on awareness. If $\in \mathcal{A}(s)$ then $[P] \subseteq \mathcal{A}(s)$. This shows that if an agent is aware of a formula then the agent is automatically aware all the epistemically equivalent formulas. For now the order of conjunction is the equivalence formula that must be satisfied. Thus, for any formula P, Q it holds that $Q \wedge P \rightarrow [P \wedge Q]$. This shows that conjuncts can be seen as epistemically equivalent, regardless of their order. Since the order of conjuncts is not relevant anymore the order of conjunction principle can hold. However, the order of conjunction principle can only hold when the epistemically equivalent formulas class holds.

The last property, is the difference between agents. The awareness logic shows whether or not an agent is aware of a formula. The awareness set could portray the difference in reasoning capacity between agents, since the awareness set is different for each agent. The examples for this principle in the previous chapter could show this: a professor is more likely to have a higher reasoning capacity than his student. This can be

shown in awareness logic; the professor is likely to be aware of more formulas since the professor has a higher reasoning capacity and can explicitly know more than his students. The student is aware of fewer formulas and thus explicitly know less than his professor. So, the difference between agents can be shown with the awareness logic.

4.2 Impossible possible words

The second approach to epistemic logic this thesis will analyse is the impossible possible worlds semantics. The impossible possible worlds semantics was introduced by Jaakko Hintikka as an solution to avoid problems that occurred by using the s5 system for epistemic logic, like logical omniscience. Hintikka states in his article *Impossible Possible Worlds Vindicated* [**1979**] that in order to solve logical omniscience, the assumption that every epistemically possible world is logically possible, must be given up. In other words, he implies that there must also be worlds that are epistemically possible but not logically possible. The idea is to enlarge the possible worlds semantics with impossible worlds. In these impossible worlds the traditional rules of logic do not hold [**Fagin et al.**, **2003**]. For example, $\varphi \wedge \neg\varphi$ can be true in an impossible world but never in a possible world. This difference exists because agents may consider these logically impossible worlds as possible worlds [**Fagin et al.**, **2003**].

The impossible worlds structure shows the formal semantics, the impossible worlds structure M, M is a tuple $(S, W, \sigma, \mathcal{K}_1, ..., \mathcal{K}_n)$. Where $(S, \mathcal{K}_1, ..., \mathcal{K}_n)$ is a Kripke frame, $W \subseteq S$ is the set of possible states or worlds and σ is a syntactic assignment which assigns truth values to all formulas in all states [**Fagin et al.**, **2003**]. σ behaves standard for the possible words, this means it follows the rules of logic. However, σ behaves arbitrary for impossible worlds, because the truths values are dependent of what an agent considers as possible. The impossible worlds allow for logical properties to change. This leads to a different examination of formulas in the impossible world. Every formula in an impossible world is evaluated arbitrarily, similar to the way atomic propositions are always evaluated. For example, φ is examined the same as in a possible world, but $\varphi \wedge \psi$ is not examined as an conjunction would be examined in a possible world.

Logical omniscience does not hold in the impossible possible worlds semantics. Because agents consider the impossible states when determining their knowledge, but the impossible states are not considered when determining logical implication [**Fagin et al.**, **2003**]. For example, an agent knows all formulas in P and P logically implies Q. Since the agent knows all formulas in P, all formulas in P must be true in all the states that the agent considers epistemically possible. But in an impossible state Q might fail to be true while P is true. Thus, the agent does not necessarily need to know Q, since Q may be false in the impossible worlds that the agent considers as possible.

An other example where logical omniscience does not hold in the impossible world semantics, is that an agent is assumed to know all tautologies in the S5 system [**Fagin et al.**, **2003**]. A tautology must be true at every possible world [**Halpern et al.**, **2011**], so when only possible worlds are examined the agent will be seen as omniscient since the agent must know all tautologies. However, when impossible worlds are introduced, the agent does not necessarily know all tautologies, since tautologies do not have to hold in impossible worlds.

4.2.1 Examples applied

The examples of chapter 3 are again used to see how well the impossible possible worlds semantics represents knowledge. The first property is factivity, the factivity principle holds when knowledge implies truth. As stated before this is also known as the reflexive principle, $K\varphi \rightarrow \varphi$. For factivity to hold in the impossible possible worlds semantics, the reflexive principle needs to hold. This means that for factivity to hold in impossible possible worlds semantics, the impossible possible worlds semantics must first satisfy the reflexivity principle. Thus, every possible world, must have a reflexive relation. The impossible worlds do not have to be reflexive since these worlds are not considered during the evaluation. However, this is not always the case. So, the impossible possible worlds semantics does not necessarily satisfy the factivity property of chapter 3.

The second property is the order of conjunction. The impossible possible worlds semantics does not satisfy the order of conjunction principle, while the possible worlds semantics does satisfy this principle. In the impossible world semantics inconsistent formulas can be true. This is because the formulas in impossible worlds can be inconsistent and logically impossible. For example, it may be the case that $\varphi \wedge \psi$ is true at an impossible world but $\psi \wedge \varphi$ is not true in that same impossible world. This shows that the order of conjunction does matter in the impossible possible world semantics. As explained above, every formula in an impossible world is examined arbitrarily. This means that $\varphi \wedge \psi$ is evaluated as a complete different formula than $\psi \wedge \varphi$. Thus the order of conjunction principle is not satisfied for impossible possible worlds semantics.

However, just as with the awareness logic, the order of conjunction principle can hold when the epistemically equivalent formulas class holds. In the impossible possible worlds semantics this class puts a constraint on the evaluation in the impossible worlds. If $P \in \sigma(w)$ for any impossible world w, then $[P] \subseteq \sigma(w)$. If an formula is true at an impossible world then his epistemic equivalent formulas are also true in the same impossible world. Again the equivalence formula that needs to hold is the order of conjunction. For any formula P, Q it holds that $Q \wedge P \rightarrow [P \wedge Q]$. So conjuncts can be seen as epistemically equivalent, regardless of their order. This means that the impossible possible worlds semantics satisfies the order of conjunction principle only if the epistemically equivalent formulas class holds.

The last property is the difference between agents in reasoning capacity. The impossible possible worlds semantics constructs impossible worlds for formulas which an agent considers as possible but are logically impossible or inconsistent. When an agent considers an other world as possible, this world is added to the model and the set of worlds. Since different agents consider different worlds as possible, the set of worlds an agent considers as possible is dependent of the agent. This means that the difference between agents can be seen in the impossible possible worlds semantics. The same example of the professor and his student can be used here. A professor is more likely to have more reasoning capacity than his student [**Hogan et al., 2001**]. Thus the professor will consider other worlds as possible than his student. The difference between agents is visible when looking at which worlds agents consider as possible. The difference between these worlds also show the difference between agents.

4.3 Awareness logic and Impossible possible worlds compared

Both Awareness logic and impossible possible worlds have been analysed in the sections above using the example properties from chapter 3. Awareness logic and the impossible

possible worlds semantics are different in their approach to improve epistemic logic. The biggest difference between the two approaches is which element of the S5 system they change. The awareness logic changes the truth evaluation of a formula. The normal evaluation for whether a formula is true or false does not hold anymore under awareness logic. Since knowledge is dependent of the awareness set of an agent. Because an agent can only explicitly know a formula if the agent knows the formula implicitly and if the agent is aware of the formula. This means a formula must always be in the awareness set of an agent in order for the agent to explicitly know that formula. The impossible possible worlds semantics keeps the truth evaluation of formulas the same as in classical logic, but the system allows for worlds to be included where logical properties can change. The formulas in impossible worlds do not have to be logically possible and, as stated in the previous section, every formula is examined arbitrarily. This means that the logical properties can change or they do not even hold their original meaning.

Besides the difference between the awareness logic and the impossible possible worlds semantics, the outcome of checking whether or not the approaches satisfy the properties from chapter 3 was the exact same. Firstly, both the awareness logic and the impossible possible worlds semantics did not necessarily satisfied factivity, in both approaches the reflexive principle must hold in order for factivity to be satisfied. Secondly, both approaches did not necessarily satisfy the order of conjuction principle. Because in both systems this principle was dependent on what an agent is aware of or considers as possible. But when the epistemically equivalent formulas hold both of the approaches did satisfy the order of conjunction principle. However, the order of conjunction is only one example of an equivalence formula, with this constraint many more examples can be introduced. Lastly, both approaches did show the difference between agents in their systems. With the awareness logic the awareness set is dependent on the agent and with the impossible possible worlds semantics the worlds the agent considers possible is dependent on the agent.

The question this chapter tries to answer was, which formal system represents knowledge more beneficial, keeping the properties of chapter 3 in mind? The awareness logic or the impossible possible worlds semantics? Because both approaches satisfied exactly the same properties from chapter 3, it is hard to tell which approach represents knowledge better than the other approach. However, both approach change a different aspect of epistemic logic. As already explained, the awareness logic changes the evaluation of a formula and the impossible possible worlds semantics changes the examination of a formula. The awareness logic feels intuitively less correct, because to change the whole evaluation of a formula in order to avoid logical omniscience seems to be going to far. The impossible possible worlds also does not feel intuitively correct since the formulas that are true in the impossible worlds can be logically incorrect and very abstract. But this is the same for the formulas in the awareness structure. An other difference between the two approaches is when they want to satisfy factivity, in both worlds the reflexive relation needs to hold. But in the impossible possible worlds semantics the the reflexive relation only needs to hold in the possible worlds and not in the impossible worlds. While in the awareness logic all of the worlds need to satisfy the reflexive relation. When the two approaches want to satisfy the order of conjunction The question which of the two approaches represents knowledge more beneficial is still to be answered, since both of the approaches performed exactly alike.

5 Conclusion

The first aim of this thesis was to analyse the S5 system, the standard system for epistemic logic by using *Dynamic epistemic logic* [**van Ditmarsch et al.**, **2007**]. However, while analysing the S5 system, some serious problems came up. For example, the S5 system requires its agents to be logical omniscient which is not the case with humans or machines [**Halpern et al.**, **2011**]. These problems show that the S5 system holds too strong restrictions on knowledge. In order to mitigate those problems a new approach to epistemic logic should be constructed. This approach should exist somewhere on the spectrum between too strong restrictions and a representation of knowledge with too little restrictions. In order to think in the correct direction for this approach, this thesis aimed to suggest three simple properties that might be beneficial for an epistemic logic. Namely, factivity, the order of conjunction, and the difference between agents. These properties were tested with two formal approaches to the epistemic knowledge: awareness logic [**Fagin et al.**, **1987**] and impossible possible worlds semantics [**Hintikka**, **1969**]. The approaches were checked to see if they satisfied the properties. Lastly, this thesis aimed to see if one of these approached represented knowledge more beneficial, in other words if one of the approaches satisfied more properties. However, the awareness logic and the impossible possible worlds semantics performed exactly similar.

All the sub-questions of this thesis have been answered, now the research question of this thesis can be answered: "What are the desired results for constructing an epistemic logic for bounded agents?". In this thesis examples for properties that an epistemic logic would want to implement were suggested, namely factivity, order of conjunction, and the difference between agents. These were very basic principles that would lead to more desired results for an epistemic logic for bounded agents. However, when the properties were tested to see if awareness logic and the impossible possible worlds semantics satisfied these properties, only one of them was satisfied. Nevertheless, the desired result of an epistemic logic is to satisfy all of the three properties. As already argued in chapter 3, factivity seems evident to implement [**Ichikawa et al.**, **2018**], knowledge should imply truth. The order of conjunction is also a principle that seems clear on why it is beneficial to implement. Since the order of formulas in a conjunction does not change the truth value in propositional logic, it also should not be changed in epistemic logic. Besides this, the order of conjunction is an equivalence principle. The difference between agents is already present in awareness logic and the impossible possible worlds semantics. This property will make a clear distinction between agents and makes it easier to reason about an other agents knowledge. So, the awareness logic and the impossible possible worlds semantics do not have the desired results yet. If both of the approaches added the potential fixes to satisfy factivity and the conjunction order, then they would have desired results. However, as mentioned, this is only one example of how to use a more general idea about adding constraints to these approaches. The way these constraints are added feels unnatural, since they are only added to force the result that is wanted. The question remains if any of the awareness logic or the impossible possible worlds semantics represent knowledge beneficial.

6 Discussion

As stated in the conclusion, the desired result for an epistemic logic for bounded agents would be to satisfy all three of the suggested properties: factivity, order of conjunction and the difference between agents. This conclusion is a stepping stone towards the construction for an epistemic logic for bounded agents. This thesis has shown that the awareness logic and the impossible possible worlds semantics are good starting points in improving the epistemic logic for bounded agents. But there have to be adaptations for them to satisfy factivity and the order of conjunction. These adaptions are to make sure that the approaches satisfy factivity and the order of conjunction. However, when applying these adaptations, it feels like they are forcing the results that seems beneficial. This does not feel natural and thus this may not be the best approach to construct an epistemic logic for bounded agents.

Of course, this thesis is only a push in the direction to construct an epistemic logic for bounded agents. It requires further research to examine more closely the link between the suggested properties and the desired results of an epistemic logic for bounded agents. Further research could be done in analysing other approaches to epistemic logic besides the awareness logic and the impossible possible worlds semantics. It would be interesting to see how other approaches handle the suggested properties. Next to this, this thesis had only time and space to suggest three different properties, there could be many other properties that would be beneficial to an epistemic logic for bounded agents. Again this requires further research. Besides further research in other approaches or other properties, the way humans perceive and think about knowledge also requires further research. It is important to know how humans think about knowledge if it is to be implemented to Artificial intelligence. Because Artificial Intelligence needs to reason about human knowledge.

As already said the findings of this thesis contribute to research about logical omniscience and the construction of an epistemic logic for bounded agents. It criticizes existing solutions to the logical omniscience problem and suggests properties that are beneficial to implement in an epistemic logic for bounded agents. It is a small step in the direction for a better knowledge representation for bounded agents. These findings will be useful when human knowledge is to be implemented to Artificial Intelligence.

References

[1] Alechina, N., Logan, B. (2002, July). *Ascribing beliefs to resource bounded agents.* In Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2 (pp. 881-888).

[2] Van Ditmarsch, H., van Der Hoek, W., Kooi, B. (2007). *Dynamic epistemic logic (Vol. 337).* Springer Science Business Media.

[3] Fagin, R., Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. Artificial intelligence, 34(1), 39-76.

[4] Fagin, R., Moses, Y., Halpern, J. Y., Vardi, M. Y. (2003). Reasoning about knowledge. MIT press.

[5] Halpern, J. Y., Pucella, R. (2011). Dealing with logical omniscience: Expressiveness and pragmatics. Artificial intelligence, 175(1), 220-235.

[6] Hogan, K., Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 38(6), 663-687.

[7] Holliday, W. H. (2015). Epistemic closure and epistemic logic I: Relevant alternatives and subjunctivism. Journal of Philosophical Logic, 44(1), 1-62.

[8] Hintikka, J. (1962) [2005], Knowledge and Belief: An Introduction to the Logic of the Two Notions, second edition, Vincent F. Hendriks and John Symons (eds.), (Texts in Philosophy, 1), London: College Publications.

[9] Hintikka, J. (1979). Impossible possible worlds vindicated. In Game-theoretical semantics (pp. 367-379). Springer, Dordrecht.

[10] Jago, M.(2006) *Hintikka and Cresswell on logical omniscience* Logic and Logical Philosophy, 15(4), 325-354.

[11] Ichikawa, Jonathan Jenkins and Matthias Steup, "The Analysis of Knowledge", The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), Consulted on October 2020 URL = ¡https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/¿.

[12] Kvanvig, J. L. (2006). Closure principles. Philosophy Compass, 1(3), 256-267.

[13] Levesque, H. J. (1984, August). *A logic of implicit and explicit belief.* In AAAI (pp. 198-202).

[14] van Linder, B., van der Hoek, W., Meyer, J. J. C. (1998). Formalising abilities and opportunities of agents. Fundamenta Informaticae, 34(1, 2), 53-101.

[15] Meyer, J. J. C., Van Der Hoek, W. (2004). Epistemic logic for AI and computer science (Vol. 41). Cambridge University Press.

[16] Rendsvig, Rasmus and John Symons, "Epistemic Logic", The Stanford Encyclopedia of Philosophy (Summer 2019 Edition), Edward N. Zalta (ed.), Consulted on September 2020 URL = ¡https://plato.stanford.edu/archives/sum2019/entries/logic-epistemic/¿

[17] Sim, K. M. (1997). *Epistemic logic and logical omniscience: A survey.* International Journal of Intelligent Systems, 12(1), 57-81.

[18] Turri, J. (2011). Mythology of the factive. Logos Episteme, 2(1), 141-150.