Chapter 2

Epistemic logic

2.1 Introduction

In Knowledge and Belief Jaakko Hintikka (1962), for the first time, described knowledge in terms of possible worlds. This research area has become known as epistemic logic. It stems from the Greek word for knowledge: ἐπιστήμη. In this thesis I take the term “epistemic” broader, applying it to belief and other ways an agent might have information as well. This concurs with much of the literature in this area. In this chapter I give an outline of epistemic logic. The notions that are explained in this chapter are used throughout this thesis. This chapter is a rather brief introduction to the subject. For a more extensive introduction see Fagin, Halpern, Moses, and Vardi (1995) or Meyer and Van der Hoek (1995).

2.2 Language and semantics

Epistemic logic can be used to model the information agents have about the world, but it is especially suited to model the information agents have about each other’s information. Suppose for example that there is a situation involving two agents, a and b. If the proposition ‘p’ means ‘it is raining’, then ‘a knows it is raining’ can be formalized as ‘\(\Box_a p\)’. The subscript indicates that we are concerned with a’s knowledge. The sentence ‘b knows that a knows it is raining’ can be formalized as ‘\(\Box_b \Box_a p\)’.

Definition 2.1 (Language of epistemic logic \(\mathcal{L}_{\mathcal{P,A}}\))

Let a countable set of propositional variables \(\mathcal{P}\) and a finite set of agents \(\mathcal{A}\) be given. The language of epistemic logic \(\mathcal{L}_{\mathcal{P,A}}\) is given by the following rule in Backus-Naur Form (BNF):

\[
\varphi ::= \bot \mid p \mid \neg \varphi \mid (\varphi_1 \land \varphi_2) \mid \Box_a \varphi
\]
where \( p \in \mathcal{P} \) and \( a \in \mathcal{A} \). Moreover \( \top \) is an abbreviation for \( \neg \bot \), \( (\varphi \lor \psi) \) is an abbreviation for \( \neg(\neg \varphi \land \neg \psi) \), \( (\varphi \rightarrow \psi) \) is an abbreviation for \( \neg \varphi \lor \psi \) and \( (\varphi \leftrightarrow \psi) \) is an abbreviation for \( ((\varphi \lor \psi) \land (\psi \lor \varphi)) \). \( \Box_a \varphi \) is an abbreviation for \( \neg \Box_a \neg \varphi \). I will also use the convention to omit the outermost parentheses of a sentence. \( \square \)

Epistemic logic is a modal logic. The standard semantics for epistemic logic does not differ much from standard semantics for modal logic. The only difference is that there is an accessibility relation for every agent.

**Definition 2.2 (Epistemic models)**

An epistemic model for \( \mathcal{L}_{\mathcal{P}, \mathcal{A}} \) is triple \( M = (W, R, V) \) such that:

- \( W \neq \emptyset \); a set of states or possible worlds;
- \( R : \mathcal{A} \rightarrow 2^{W \times W} \); assigns an accessibility relation to each agent;
- \( V : \mathcal{P} \rightarrow 2^W \); assigns a set of possible worlds to each propositional variable.

If \( M = (W, R, V) \) is a model, a pair \( (M, w) \), where \( w \in W \) is called a pointed model. A pair \( F = (W, R) \) is called a frame and a pair \( (F, w) \), where \( w \in W \), is called a pointed frame. A pair \( (F, V) \) is also a model. But I will be somewhat sloppy with the terminology. \( \square \)

The class of all frames is called \( K_\mathcal{A} \). The class of all pointed frames is called \( K_{\mathcal{P}, \mathcal{A}} \). The class of all models for \( \mathcal{L}_{\mathcal{P}, \mathcal{A}} \) is called \( \mathcal{K}_{\mathcal{P}, \mathcal{A}} \), and \( \mathcal{K}_{\mathcal{P}^*, \mathcal{A}} \) is the class of all pointed models for that language.

**Definition 2.3 (Semantics for \( \mathcal{L}_{\mathcal{P}, \mathcal{A}} \))**

Let an epistemic model \( (M, w) \) where \( M = (W, R, V) \) be given. Let \( p \in \mathcal{P} \), \( a \in \mathcal{A} \), and \( \varphi, \psi \in \mathcal{L}_{\mathcal{P}, \mathcal{A}} \).

\[
\begin{align*}
(M, w) \not\models \bot & \quad \text{iff} \quad w \notin V(p) \\
(M, w) \models p & \quad \text{iff} \quad (M, w) \not\models \neg p \\
(M, w) \models (\varphi \land \psi) & \quad \text{iff} \quad (M, w) \models \varphi \text{ and } (M, w) \models \psi \\
(M, w) \models \Box_a \varphi & \quad \text{iff} \quad (M, v) \models \varphi \text{ for all } v \text{ such that } w R(a) v
\end{align*}
\]

where \( w R(a) v \) is an abbreviation for \( (w, v) \in R(a) \). \( \square \)

Consider the following example. Suppose two children, \( a \) and \( b \), have been playing outside. Both of them can see whether the other child’s face is muddy, yet they cannot see their own faces. This situation can be analyzed using an epistemic logic by making a model for this situation. A picture of this model is shown in figure 2.1. The states are indicated by pairs \( (x, y) \), where \( x \) and \( y \) stand for the state \( a \)’s respectively \( b \)’s face is in. The number \( 0 \) means the child’s face is not
muddy, 1 means it is muddy. The solid and dashed lines represent the accessibility relations of \( a \) and \( b \) respectively. In \((0, 1)\) for example, \( a \) cannot rule out her face is muddy, but she also cannot rule out she is not muddy (which is actually the case). Therefore \((0, 1)\) and \((1, 1)\) are both accessible to \( a \). She does know that \( b \)'s face is muddy. Therefore in both worlds that are accessible to her \( b \)'s face is muddy. This example is a specific instance of the initial situation of the muddy children puzzle, which is discussed in chapter 4 (page 35).

These so-called Kripke models have an important property. They do not only model what information the agents have about the world, they also model ‘higher-order information’, i.e. the information the agents have about the information that the agents have, and so on. For example, in the Kripke model above, \( a \) knows that \( b \) does not know whether she is muddy or not. But it does not stop there, because the model also gives us that \( b \) knows that \( a \) does not know that \( b \) knows that \( a \) is muddy. In this way the model allows us to stack these kinds of constructions indefinitely, and thus it models all higher-order information at once.

For the notion of validity we overload \( \models \) with the following notions.

**Definition 2.4 (Validities)**

Let \( M = (W, R, V) \) be a model, \( F = (W, R) \) be a frame \( w \in W \) be a world. Let \( S_{P, A} \) and \( *S_{P, A} \) be classes of models and pointed models respectively.

\[
\begin{align*}
\models_{(M, w)} \varphi & \iff (M, w) \models \varphi \\
\models s_{P, A} \varphi & \iff \models_{(M, w)} \varphi \text{ for every } (M, w) \in s_{P, A} \\
\models M \varphi & \iff \models_{(M, w)} \varphi \text{ for every } w \in W \\
\models s_{P, A} \varphi & \iff \models_{M} \varphi \text{ for every } M \in s_{P, A} \\
\models_{(F, w)} \varphi & \iff \models_{(F, V, w)} \varphi \text{ for every } V : P \to 2^W \\
\models s_{A} \varphi & \iff \models_{(F, w)} \varphi \text{ for every } (F, w) \in s_{A} \\
\models F \varphi & \iff \models_{(F, w)} \varphi \text{ for every } w \in W \\
\models s_{A} \varphi & \iff \models_{F} \varphi \text{ for every } F \in s_{A}
\end{align*}
\]

Generally by \( \models \varphi \) we mean \( \models s_{P, A} \varphi \), and by \( \Gamma \models \varphi \) we mean local logical
consequence, i.e. for every pointed model \((M, w) \in \mathcal{S}_{\mathcal{P}, \mathcal{A}}\), if \((M, w) \vDash \psi\) for every \(\psi \in \Gamma\), then \((M, w) \vDash \varphi\).

Whether these notions coincide depends on the choice of models, frames, and classes of these.

### 2.3 Proof systems

The simplest proof system for epistemic logic is \(K_{\mathcal{P}, \mathcal{A}}\).

**Definition 2.5 (Proof system \(K_{\mathcal{P}, \mathcal{A}}\))**

Let \(\varphi, \psi\) be sentences in \(\mathcal{L}_{\mathcal{P}, \mathcal{A}}\) and let \(a\) be an agent in \(\mathcal{A}\). The proof system \(K_{\mathcal{P}, \mathcal{A}}\) consists of the following axioms and derivation rules.

- **Taut** all instantiations of propositional tautologies
- **Distr** \(\Box_a(\varphi \rightarrow \psi) \rightarrow (\Box_a \varphi \rightarrow \Box_a \psi)\) (distribution)
- **MP** \(\varphi \rightarrow \psi\)
- **Nec** \(\varphi \rightarrow \Box_a \varphi\) (necessitation)

Let us now introduce a general notion of provability that can be used for other proof systems as well.

**Definition 2.6 (Provability)**

Let \(S\) be a proof system and \(\mathcal{L}\) a logical language. A derivation or proof in \(S\) consists of a sequence of sentences of \(\mathcal{L}\) each of which is an instance of an axiom or is the result of applying a derivation rule to sentences that occur earlier in the sequence. If \(\varphi\) is the last sentence in a derivation, then \(\varphi\) is provable, or deducible, in \(S\), notation \(\vdash_S \varphi\).

Note that this notion of provability excludes the possibility of premises of an inference. The rules **MP** and **Nec** can only be applied when the formulas above the line have already been deduced.

The system \(K_{\mathcal{P}, \mathcal{A}}\) is sound and complete with respect to the class of models \(K_{\mathcal{P}, \mathcal{A}}\). That is the notion of validity and deducibility coincide.

**Theorem 2.1 (Soundness and completeness of \(K_{\mathcal{P}, \mathcal{A}}\))**

\[ \vdash_{\mathcal{P}, \mathcal{A}} \varphi \text{ iff } \models_{\mathcal{P}, \mathcal{A}} \varphi \]

for every sentence \(\varphi \in \mathcal{L}_{\mathcal{P}, \mathcal{A}}\). The if above is called completeness, the only if is called soundness.
2.4. General and common knowledge

Soundness is often easy to prove by induction on the length of the proof. Completeness is usually shown by contraposition and involves the construction of a canonical model. The notion of completeness presented here is called weak completeness. For strong completeness, see chapter 3.

Epistemic logic is used for many notions involving information. The axioms and rules of $K_{p,A}$ are minimal requirements these notions should meet. For many epistemic notions there are more requirements that should be met. For example in the case of knowledge, you want that if an agent knows something, then it is true:

$$\text{T} \quad \Box \phi \rightarrow \phi \quad \text{(factivity)}$$

There are two systems that are especially relevant for epistemic logic, viz. $KD_{45,p,A}$ and $S_{5,p,A}$.

1. The system $KD_{45,p,A}$ is usually considered to be the best system to model belief, and $S_{5,p,A}$ is considered to be the best system to model knowledge.

2. $D$ \hfill $\Box \phi \rightarrow \Diamond \Diamond \phi$ \hfill (positive introspection)

3. $\Box \phi \rightarrow \Box \Diamond \phi$ \hfill (positive introspection)

The system $S_{5,p,A}$ consists of all axioms and rules of $K_{p,A}$ and the axioms T, 4, and 5 for sentences of $L_{p,A}$. The system $KD_{45,p,A}$ consists of all axioms and rules of $K_{p,A}$ and the axioms D, 4, and 5 for sentences of $L_{p,A}$.

By a class of pointed models $\star S_{p,A}$ that is intended to be complete with respect to a proof system $S$, the following is meant. Take the subclass of frames $\star S_A$ such that all the axioms and the rules of $S$ are valid in those models. Then take the class of pointed models $\star S_{p,A}$ associated with $\star S_A$. It turns out that the axioms of $KD_{45,p,A}$ are valid on the class of serial, transitive, and euclidean frames. The axioms of $S_{5,p,A}$ are valid on those frames where the accessibility relations are all equivalence relations.

2.4 General and common knowledge

There are two more important concepts in epistemic logic: general knowledge and common knowledge. If something is general knowledge, it means that everybody knows it. Common knowledge is a typical concept for epistemic logic. It is concerned with information about information. If something is common knowledge, then it is general knowledge, but it is also general knowledge that it is general

---

1. The systematic name of $S_{5,p,A}$ would be $KT_{45,p,A}$, however this name is not commonly used. Moreover there are more systems that would have the same set of valid sentences. For example $KT_{5,p,A}$ or $KDB_{5,p,A}$ or $KTB_{4,p,A}$. So I choose for $S_{5,p,A}$.

2. Sometimes completeness is proved with respect to a smaller class of models than those where the axioms hold. For example the logic of linear time is valid on all linear frames, but also on other frames.
knowledge, and it is also general knowledge that it is general knowledge, and so on ad infinitum. Common knowledge is a very useful concept, especially in the context of games.

To incorporate these two notions in the formal language two modal operators have to be added: $E_B \varphi$, which can be read 'every member of group $B$ knows that $\varphi$' and $C_B \varphi$, which can be read 'it is common knowledge among members of $B$ that $\varphi$'. Let $\mathcal{L}^{E\!C}_{P,A}$ be the language extended with these two operators. The models do not have to be changed or adapted to interpret sentences in which these operators occur.

**Definition 2.7 (Semantics for $E$ and $C$)**

Let a model $(M, w)$ where $M = (W, R, V)$ be given. Let $p \in \mathcal{P}$, $\varphi \in \mathcal{L}^{E\!C}_{P,A}$, and $B \subseteq A$.

$$(M, w) \models E_B \varphi \iff (M, v) \models \varphi \text{ for all } v \text{ such that } w R(B) v$$

$$(M, w) \models C_B \varphi \iff (M, v) \models \varphi \text{ for all } v \text{ such that } w R(B)^+ v$$

where $R(B) = \bigcup_{a \in B} R(a)$, and $R(B)^+$ is the transitive closure of this relation.

In this definition the accessibility relation corresponding to the common knowledge operator is interpreted as the transitive closure of the union of the accessibility relations of the agents in the group. In other approaches it is taken to be the reflexive transitive closure of this relation. However, this is not practical when other epistemic notions such as belief are studied. Nevertheless, when we are working within the class $S5_{P,A}$ it is the reflexive transitive closure.

For the proof system, we need two additional axioms and an additional rule.

- **E** $E_B \varphi \leftrightarrow \bigwedge_{a \in B} \Box a \varphi$
- **Mix** $C_B \varphi \rightarrow E_B (\varphi \land C_B \varphi)$
- **Ind** $\varphi \rightarrow E_B (\psi \land \varphi)$

The system $K^{E\!C}_{P,A}$ consist of all axioms and rules of $K_{P,A}$ and the axioms **E**, **Mix**, and rule **Ind** for sentences of $\mathcal{L}^{E\!C}_{P,A}$. And in the same fashion we get $KD45^{E\!C}_{P,A}$ and $S5^{E\!C}_{P,A}$. Again these systems are sound and complete with respect to $*K_{P,A}$, $*KD45_{P,A}$, and $*S5_{P,A}$.

### 2.5 Bisimulation

A useful notion that stems from modal logic is bisimulation.

**Definition 2.8 (Bisimulation)**

Let two models $M = (W, R, V)$ and $M' = (W', R', V')$ in the class $K_{P,A}$ be given. A relation $\mathfrak{R} \subseteq W \times W'$ is a bisimulation iff for all $w \in W$ and $w' \in W'$ with $w \mathfrak{R} w'$:
2.5. Bisimulation

**Atoms** \( w \in V(p) \) iff \( w' \in V'(p) \) for all \( p \in \mathcal{P} \)

**Forth** for all \( a \in \mathcal{A} \) and all \( v \in W \), if \( w R(a) v \), then there is a \( v' \in W' \) such that
\( w' R'(a) v' \) and \( v R' v' \)

**Back** for all \( a \in \mathcal{A} \) and all \( v' \in W' \), if \( w R'(a) v' \), then there is a \( v \in W \) such that
\( w R(a) v \) and \( v R v' \)

We write \((M, w) \preceq (M', w')\), iff there is a bisimulation between \( M \) and \( M' \) linking \( w \) and \( w' \). Then we call \((M, w)\) and \((M', w')\) bisimilar. \( \square \)

The main theorem about bisimulation is:

**Theorem 2.2**
Let two models \((M, w)\) and \((M', w')\) in \( K_{\mathcal{P}, \mathcal{A}} \) be given. If \((M, w) \preceq (M', w')\), then for every \( \varphi \in \mathcal{L}_{E, \mathcal{P}, \mathcal{A}} \) it holds that \((M, w) \models \varphi\) iff \((M', w') \models \varphi\). \( \square \)

It generally does not hold vice versa. There are a number of cases when it does hold vice versa, for example if for every world the set of accessible worlds is finite.

A generalization of bisimulation will play an important role in section 6.5.