Logics for Reasoning About Uncertainty

(Draft)

Prof. Dr. Erich Grädel

SS 2017

Chapter 1

Logics of Knowledge

In the first chapter we investigate logics for reasoning about knowledge. We assume that there are multiple agents with different states of knowledge, and we want to make statements about the knowledge of the agents concerning facts about the outside world about each other's knowledge.

Let us start with a simple example.

> Three logicians come into a bar.
>
> Barkeeper: *Beer everybody?*
> First logician: *I don't know.*
> Second logician: *I don't know.*
> Third logician: *Yes!*

Obviously, for this simple example it is not difficult to figure out why the agents reason in the way the do, and how the state of their knowledge develops as the conversation proceeds. But for more involved sutuations this may be far less obvious and is useful to have a formal framework for modeling knowledge.

A popular such framework is based on *possible worlds semantics* and *modal logics.* The intuitive idea behind the possible-worlds framework is that besides the true state of affairs, there are a number of other possible states of affairds or "worlds". An agent may not be able to tell which of a number of possible states describes the actual world. He is then said to *know* a fact if this fact is true at all the worlds he considers possible (given his current state of information).

To describe these ideas more precisely, we first need an appropriate language. We find one in *modal logic.*

1.1 Modal Logics

For modeling a situation in modal logic *(ML)* we use

- a finite, non-empty set A of agents,
- a set $(P_i)_{i\in I}$ of atomic propositions,
- a *Kripke structure* $\mathcal{K} = (W, (E_a)_{a\in A}), (P_i)_{i\in I}$, where W is a set of worlds, $E_a \subseteq W \times W$, and $P_i \subseteq W$ for each $i \in I$.

For an agent $a \in A$ in world $v \in W$ of the Kripke structure $\mathcal{K}$, the set $vE_a := \{w : (v, w) \in E_a\}$ is supposed to be the set of worlds that he considers possible.

Formulae of modal logics will always be evaluated at a specific world w of a Kripke structure.

Notice that modal logics have many interpretations, and reasoning about knowledge is just one of them. If a modal logic is used as a logic of knowledge, like it is here, one usually requires that all relations E_a are reflexive, symmetric and transitive, and hence are equivalence relations on W. This captures the intuition that an agent considers world t possible in world s if in both worlds s and t he has the same information, i. e., they are indistinguishable to him.

Let us have a look again at the introductory example. Represented as a Kripke structure, we have $A = \{L_1, L_2, L_3\}$, where each L_i denotes one of the logicians. Further, we define $W = \{0,1\}^3$ and $P_i = \{(w_1, w_2, w_3) : w_i = 1\}$ such that P_i stands for L_i wanting beer. The edges are $E_{L_i} = \{((v_1, v_2, v_3), (w_1, w_2, w_3)) : v_i = w_i\}$ since each logician must be completely open about the other's decision about having a beer; the only thing he knows for sure is whether he wants one for himself.

The formulae of modal logic (ML) are defined by the grammar

$$\varphi ::= P_i \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (\varphi \wedge \varphi) \mid \langle a\rangle\varphi \mid [a]\varphi.$$

For the semantics we define

- $\mathcal{K}, w \models P_i :\Longleftrightarrow w \in P_i$,
- $\mathcal{K}, w \models \neg\varphi :\Longleftrightarrow \mathcal{K}, w \not\models \varphi$,
- $\mathcal{K}, w \models \varphi \vee \psi :\Longleftrightarrow \mathcal{K}, w \models \varphi$ or $\mathcal{K}, w \models \psi$,
- $\mathcal{K}, w \models \varphi \wedge \psi :\Longleftrightarrow \mathcal{K}, w \models \varphi$ and $\mathcal{K}, w \models \psi$,
- $\mathcal{K}, w \models \langle a\rangle\varphi :\Longleftrightarrow \mathcal{K}, z \models \varphi$ for some $z \in wE_a$,

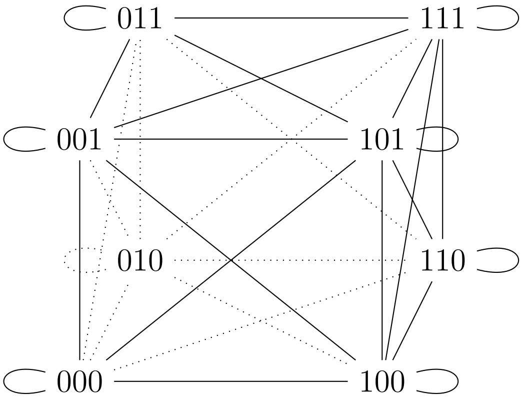


Figure 1.1: Three logicians in a bar. Equivalence classes for one actor are two opposite faces of the cube. Note that there is no optical distinction between edges of different actors.

- $\mathcal{K}, w \models [a]\varphi :\Longleftrightarrow \mathcal{K}, z \models \varphi$ for all $z \in wE_a$.

In our context of reasoning about knowledge we write $K_a\varphi$ instead of $[a]\varphi$ and say “agent *a knows* φ”. We usually will not use the notation $\langle a\rangle\varphi$ but instead $\neg K_a\neg\varphi$.

Coming back to the example above, we can reformulate the question of the barkeeper as $\varphi = P_1 \wedge P_2 \wedge P_3$. If the first logician does not want a beer, his answer would be “No” right away since then not everybody wants a beer. So the interesting case is when he answers “I don’t know”, as he does. By this he implicitly states that he himself wants a beer, which eliminates all worlds where he does not and thus changes the knowledge of his two fellow logicians. Yet the number of possible worlds is still four still (the right face of the cube), and they do not all agree on φ, as formally captured by $\mathcal{K}, (1, w_2, w_3) \models \neg K_1\varphi \wedge \neg K_1\neg\varphi$, so the second logician will not say “Yes” either. Repeating the argument, we see that the answer of the second logicians further restricts the set of possible worlds to 110 and 111, so eventually, the third logician, by knowing that he himself want a beer, can answer “Yes” to the barkeeper.

We see that from our formalization alone we did not get a direct way of how to solve the situation. Indeed, our reasoning was not completely “inside logic” but we also argued on the “meta–level”, for the reductions of the cube. This is a general phenomenon. Modal logic (even when extended by additional logical operators) is rarely enough to capture everything that is relevant in a situation. Rather, we look at it as a helpful tool which is used to somewhat illustrate the situation and make it easier to understand and reason about it.

1.2 The Muddy Children Puzzle

Let us discuss a more elaborate example.

We imagine n children playing together. After their play, k of them have mud on their forehead. Each one can see the mud on others but not on its own forehead. Now, their father comes and tells them all together that at least one of them has mud on his forehead. Further, he repeatedly asks the question "Do you know whether there is mud on your own forehead?". Assuming that all children are perceptive, honest, intelligent and answer simultaneously, what happens is the following. To the first $k-1$ questions, all children answer "No". To the k-th question however, exactly the muddy children answer "Yes".

Now, why is that? In the case of only one child having mud on his forehead, this child will answer "Yes" to the first question of the father since it can see that all the other children are clean. If there are two muddy children, both will answer "No" to the first question since no one knows the exact number of muddy children at that point of time. With the question having been asked a second time however, the muddy children can deduce that there is more than one child with mud on its forehead, since if there was only one, this one would have answered "yes" in the first round, as just explained. Hence, each child that sees only one muddy child, deduces that that there are exactly two muddy children and that it must be one of them. We can argue analogously for arbitrary k, with ever more deeply nested reasoning.

Assuming that $k > 1$, we further observe that, when the father tells his children at the very beginning, that at least one of them has mud on its forehead, he actually states a fact that is already known to all of them. Nevertheless this statement is crucial for the argumentation above to work. The reason is that even if it is true that everyone knows that at least one of them has mud on its forehead, it is not true that everyone knows that everyone knows that at least one of them has mud on his forehead. To establish this is exactly the role of the father's statement. If we look again at the case $k = 2$, we said that "the muddy children can deduce that there is more than one child with mud on his forehead, *since if there was only one, this one would have answered "yes" in the first round*". But this deduction is possible only if all the (muddy) children know that all children know that there is at least one child with mud on his forehead. Otherwise their reasoning could not refer to the case $k - 1 = 1$ for which the father's statement is actually new information for the muddy child.

As before, we can formalize this situation in modal logic. Nodes of the Kripke structure are tuples $(w_1, \ldots, w_n) \in \{0, 1\}^n$ with $w_i = 1$ if the ith child is muddy and $w_i = 0$ otherwise. If there are three children n, we again have a Kripke structure with the shape of a cube, but this time the equivalence classes are parallel edges. The reason for this is that in each world, a child considers possible precisely two worlds which are identical with respect to the other children being muddy or not, but differ with respect to that child itself being muddy or not. (contrary to the situation of three logicians in a bar).

Also, there is a similar way to use this formalization to resolve the situation. After the father has stated that at least one of the children has mud on his forehead, we would eliminate the node $(0, \dots, 0)$. After he has asked his question for the first time and no one answers "yes" (for $k > 1$), we would eliminate all nodes with exactly one 1, for reasons already made clear. After he has asked it for the second time and no one answers "yes" (for $k > 2$), we would eliminate all nodes with exactly two 1's, and so on. Eventually, after the k-th question, only tuples with at least $k+1$ 1's are left, meaning that everyone will know that there are at least $k+1$ muddy children and thus those will know about them being muddy.

1.3 Common Knowledge and Distributed Knowledge

We next introduce three operators formalizing notions that are important when reasoning about knowledge and that already have been in play in the previous examples. Let G denote any subset of the set A of agents.

- E_G: "everybody in G knows..."
- C_G: "it is common knowledge among agents in G that..."
- D_G: "it is distributed knowledge in G that..."

Example 1.3.1. The formula $K_1 C_{\{2,3\}}\varphi$ says that agent 1 knows that φ is common knowledge among agents 2 and 3.

The E_G–operator does not actually increase the expressivenes of the logic. It can simply be defined as

$$\mathcal{K}, w \models E_G\varphi :\Longleftrightarrow \mathcal{K}, w \models \bigwedge_{a \in G} K_a\varphi.$$

Example 1.3.2. In the muddy children example with $k > 1$, even before the father has said that there is at least one child with mud on his forehead, all children knew this. Formally, $E_A\neg(0, \dots, 0)$. If $k = 1$, we could only say $E_{A\setminus\{a_{muddy}\}}\neg(0, \dots, 0)$, where a_{muddy} denotes the single muddy child.

Now, we can define $E^0_G\varphi := \varphi$ and $E^{i+1}_G\varphi := E_G E^i_G\varphi$ and based on that the operator for common knowledge

$$\mathcal{K}, w \models C_G\varphi :\Longleftrightarrow \mathcal{K}, w \models E^i_G\varphi \text{ for all } i \in \mathbb{N}$$

which cannot be expressed by a formula of standard modal logic any longer. Semantically it stands for “Everybody knows ...” and “Everybody knows that everybody knows ...” and “Everybody knows that everybody knows that everybody knows ..., etc. ad infinitum.

Example 1.3.3. After the father has said “At least one of you has mud on his forehead” this is commong knowledge among all children, or formally $C_A(P_1 \vee \dots P_n)$.

Furthermore, for a Kripke structure $\mathcal{K}$, a group G of agents, a world v, we define

$$R^k_G(v) := \{w : w \text{ is reachable from } v \text{ by a path of length} \leq k \text{ in } \bigcup_{a \in G} E_a\},$$

and its closure

$$R^\infty_G(v) := \bigcup_{k \in \mathbb{N}} R^k_G(v).$$

This enables us to express the E^i_G–operator and the C_G–operator in graph–theoretical terms:

Lemma 1.3.4. 1. $\mathcal{K}, v \models E^k_G\varphi \Longleftrightarrow \mathcal{K}, w \models \varphi$ for all $w \in R^k_G(v)$.

2. $\mathcal{K}, v \models C_G\varphi \Longleftrightarrow \mathcal{K}, w \models \varphi$ for all $w \in R^\infty_G(v)$.

We also note that there is a connection to the *modal μ–calculus**, as captured by

Remark 1.3.5.

$$\mathcal{K}, v \models C_G\varphi \Longleftrightarrow \mathcal{K}, v \models \nu X.(\varphi \wedge \bigwedge_{a \in G} K_a X).$$

To say that φ is distributed knowledge among the group of agents G we use the D_G–operator which is defined as

$$\mathcal{K}, w \models D_G\varphi :\Longleftrightarrow \mathcal{K}, z \models \varphi \text{ for all } z \text{ with } (w, z) \in \bigcap_{a \in G} E_a.$$

Note that for larger G, the set $\bigcap_{a \in G} E_a$ tends to be smaller. Semantically, the D_G–operator pools the insights of all the agents in G. A fact is distributed knowledge in G if it holds at every world that is considered possible by every agent in G.

Example 1.3.6. Two agents $1, 2$ play a card game with cards A, B, C. Each agent gets one card, the third one is put face down on the table.

This creates six possible worlds, each described by a pair (X_1, X_2), where $X_i \in \{A, B, C\}$ and $X_1 \neq X_2$. Atomic propositions are $P_{i,X}$ with $i \in \{1, 2\}$ and $X \in \{A, B, C\}$, stating that agent i holds card X, i. e.,

$$\mathcal{K}, (X_1, X_2) \models P_{i,X} \iff X = X_i.$$

We want to make some statements about this situation, particularly with the new operators. Facts like $\mathcal{K}, (X_1, X_2) \models K_i P_{i,X_i} \wedge \neg K_i P_{1-i,X_{i-1}}$, which follow immediately from the rules of the game, are easy to say already. For not so obvious sentences, let us illustrate the situation first.

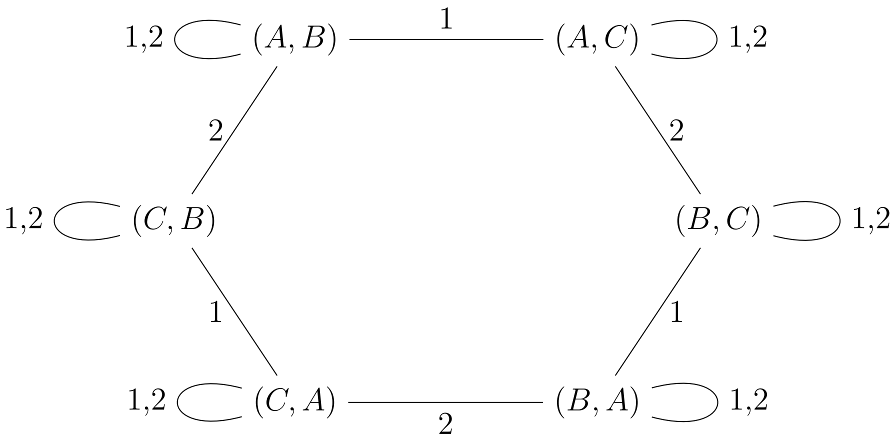


Figure 1.2: Kripke structure of a card game for two players.

Exemplarily, we can now convince ourselves of the formulae

- $\mathcal{K}, (A, B) \models K_1(P_{2,B} \vee P_{2,C}) \wedge K_1 \neg K_2 P_{1,A}$,
- $\mathcal{K}, (A, B) \models C(P_{1,A} \vee P_{1,B} \vee P_{1,C}) \wedge C(P_{1,A} \to (P_{2,B} \vee P_{2,C}))$,
  - In fact, every formula that holds at all possible worlds is common knowledge. This will be called *Generalization Rule*.
- $\mathcal{K}, (A, B) \models D(P_{1,A} \wedge P_{2,B})$.

1.4 Properties of Knowledge

So far we have described a language with modal operators such as K_a or D_G and defined a notion of truth (semantics, logically speaking) which determines whether certain facts, written as a formula in this language, are true at some possible world. We have also used it already in various examples and shown its usefulnes there.

Still, we have not argued yet whether this concept of knowledge actually captures in general our intuitive understanding of knowledge appropriately. We can attempt to do so by examining the properties of knowledge under our interpretation. One way of characterizating the properties of our interpretation of knowledge is by characterizing the formulae that are always true, i. e., that hold at every node of every Kripke structure (that speaks about knowledge). Regarding this, we will say "φ is *valid* in $\mathcal{K}$" and write $\mathcal{K} \models \varphi$ if $\mathcal{K}, w \models \varphi$ for all w in $\mathcal{K}$. Moreover, $\models \varphi$ if $\mathcal{K} \models \varphi$ for every (suitable) Kripke structure $\mathcal{K}$, and we will just say φ *is valid* in this case.

The first important property of our definition of knowledge we want to talk about is that each agent knows all the logical consequences of his knowledge. If an agent knows ψ and knows that ψ implies φ, he also knows φ. Formally speaking,

$$\models (K_a\psi \wedge K_a(\psi \to \varphi)) \to K_a\varphi.$$

This holds because for arbitrary $\mathcal{K}$ and one of its nodes w, when assuming $\mathcal{K}, w \models K_a\psi \wedge K_a(\psi \to \varphi)$ we can deduce that $\mathcal{K}, z \models \psi$ and $\mathcal{K}, z \models \psi \to \varphi$ for all $z \in wE_a$, and thus $\mathcal{K}, w \models K_a\varphi$.

We call this *Distribution Axiom (K)*. Note that it is actually an axiom *schema*, representing all formulae of this form where φ and ψ have been replaced by modal formulae, but we ussually go without this subtlety, also in the case of other axioms/axiom schemas. The axiom suggests that our agents are quite powerful reasoners. The same is suggested by the Generalization Rule (cf. Example 1.3.6; second formula), which says that every agent knows all the formulae that are valid in a given Kripke structure:

$$\text{If } \mathcal{K} \models \varphi \text{ then } \mathcal{K} \models K_a\varphi.$$

If φ is valid in $\mathcal{K}$, then φ holds at all worlds w of $\mathcal{K}$, in particular for any v it holds at all worlds w that agent a considers possible at v. Hence, $\mathcal{K}, v \models K_a\varphi$. It follows that $\mathcal{K} \models K_a\varphi$.

Note that this is not the same as saying $\mathcal{K} \models \varphi \to K_a\varphi$ which is not always true. An agent does not necessarily know all things that are true. For example, in the case of the muddy children, it may be true that some child has a muddy forehead but he might not know this. However, agents do know all *valid* formulae. Intuitively, these are the formulae that are *necessarily* true, as opposed to the formulae that just happen to be true at a given world. Being able to determine all the valid formulae is what makes our agents powerful reasoners once more.

Although an agent may not know facts that are true, it is the case that if he knows a fact, then it is true. Or put the other way around, an agent cannot know what is not true:

$$\models K_a\varphi \to \varphi.$$

We call this property the *Knowledge Axiom (T)*. It follows because the current world of an agent is always one of the worlds that he considers possible. So if $K_a\varphi$ holds at a particular world $\mathcal{K}, v$, then φ is true at all worlds agent a considers possible, in particular at $\mathcal{K}, v$. Here, we basically use that E_a is assumed to be reflexive. In contrast, for the Distribution Axiom and the Generalization Rule we did not rely on any assumptions about the edge relations at all.

The last two properties we want to consider say that agents can do introspection with respect to their knowledge. They know what they know and what they do not know:

$$\models K_a\psi \to K_aK_a\psi,$$

$$\models \neg K_a\psi \to K_a\neg K_a\psi.$$

The first one we call *Positive Introspection Axiom (4)* and the second one *Negative Introspection Axiom (5)*. Like the Knowledge Axiom they do not hold for Kripke structures in general but require certain assumptions on the edge relations. For (4) E_a has to be transitive, and for (5) E_a has to be *Euclidian*, i. e., $(v, w) \in E_a$ and $(v, z) \in E_a$ imply $(w, z) \in E_a$.

Before we will prove these axioms we want to state a few general properties of some restrictions one can place on binary relations E_a in

Lemma 1.4.1. 1. If E_a is reflexive and Euclidian, then E_a is symmetric and transitive.

2. If E_a is symmetric and transitive, then E_a is Euclidian.

3. TFAE:

   - E_a is an equivalence relation.
   - E_a is symmetric, transitive and *serial*, the latter being that for each v there is a w with $(v, w) \in E_a$.
   - E_a is reflexive and Euclidian.

Proof. 1. – $E_avw \overset{refl.}{\Longrightarrow} E_avv \wedge E_avw \overset{Eucl.}{\Longrightarrow} E_awv$.

– $E_avw \wedge E_awz \overset{sym.}{\Longrightarrow} E_awv \wedge E_awz \overset{Eucl.}{\Longrightarrow} E_avz$.

2. $E_avw \wedge E_avz \overset{sym.}{\Longrightarrow} E_awv \wedge E_avz \overset{trans.}{\Longrightarrow} E_awz$.

3. Follows with 1. and 2.

□

Other knowledge axioms that are valid in Kripke structures with specific properties only, are the so–called *Consistency Axiom (D)*, that is $\neg K_a0$, or the axiom $\varphi \rightarrow K_a\neg K_a\neg\varphi$.

Theorem 1.4.2. 1. If E_a is reflexive, then $\mathcal{K} \models K_a\varphi \rightarrow \varphi$ (T).

2. If E_a is transitive, then $\mathcal{K} \models K_a\varphi \rightarrow K_aK_a\varphi$ (4).

3. If E_a is Euclidian, then $\mathcal{K} \models \neg K_a\varphi \rightarrow K_a\neg K_a\varphi$ (5).

4. If E_a is serial, then $K \models \neg K_a0$ (D).

5. If E_a is symmetric, then $\mathcal{K} \models \varphi \rightarrow K_a\neg K_a\neg\varphi$.

Proof. 1. If $\mathcal{K}, v \models K_a\varphi$, then by reflexivness $v \in vE_a$ and hence $\mathcal{K}, v \models \varphi$. Thus, $\mathcal{K}, v \models K_a\varphi \rightarrow \varphi$ for all v.

2. If $\mathcal{K}, v \models K_a\varphi$ and $(v, w) \in E_a$ and $(w, z) \in E_a$, then $(v, z) \in E_a$ and hence $\mathcal{K}, z \models \varphi$. So $\mathcal{K}, v \models K_a\varphi \rightarrow K_aK_a\varphi$ for all v.

3. Suppose that $\mathcal{K}, v \models \neg K_a\varphi$, hence there exists $z \in vE_a$ with $\mathcal{K}, z \models \neg\varphi$. To prove that $\mathcal{K}, v \models K_a\neg K_a\varphi$ we have to show that for all $w \in vE_a$ there exists some $v \in wE_a$ with $\mathcal{K}, v \models \neg\varphi$. Since E_a is Euclidian, $z \in vE_a$ and $w \in vE_a$ imply that $z \in wE_a$. Set $v := z$.

4. $\mathcal{K}, v \models K_a0$ holds if, and only if, $vE_a = \emptyset$.

5. Assume $\mathcal{K}, v \models \varphi$. To prove that $\mathcal{K}, v \models K_a\neg K_a\neg\varphi$ we have to show that for all $w \in vE_a$ there is $z \in wE_a$ with $\mathcal{K}, z \models \varphi$. By symmetry we can set $z := v$.

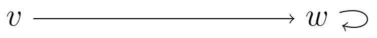
□

In particular, with Lemma 1.4.1 it follows that all these axioms hold if E_a is an equivalence relation; they are valid in all Kripke structures that speak about knowledge.

Now that we have established several logical consequences of graph properties, we ask ourselves whether their converse also holds, so for example whether for each Kripke structure $\mathcal{K}$ with $\mathcal{K} \models K_a\varphi \rightarrow \varphi$ it is reflexive. The answer is no, at least not in this direct way. As a matter of fact, there exist Kripke structures which are models of all axioms from Theorem 1.4.2

but E_a is not even reflexive, like the following one with atomic proposition $P = \{v, w\}$.

$\mathcal{K}$:



$\mathcal{K}$ is obviously not reflexive. To see that it satisfies the axioms from Theorem 1.4.2 is straightforward after one has observed that $\mathcal{K}, v \equiv_{\mathrm{ML}} \mathcal{K}, w$ (i. e., for every modal formula φ we have that $\mathcal{K}, v \models \varphi \Leftrightarrow \mathcal{K}, w \models \varphi$), provable by a simple induction on the structure of modal formulae or directly following from that fact that v and w are *bisimulation invariant**.

We are only able to establish a correspondence in both directions between the axiom systems and graph properties if we generalize our notion of a Kripke structure to that of a *Kripke frame.*

Definition 1.4.3. For a finite set A of agents, a Kripke frame is a structure $\mathcal{F} = (W, (E_a)_{a \in A})$. A Kripke structure is *based on* $\mathcal{F}$ if it expands the frame by interpretations $P_i \subseteq W$.

Definition 1.4.4. A class $\mathcal{C}$ of frames is characterized by a set Φ of modal axioms if for every Kripke frame $\mathcal{F}$ we have that $\mathcal{F} \in \mathcal{C}$ if, and only if, $\mathcal{K} \models \Phi$ for all (suitable) Kripke structures based on $\mathcal{F}$.

With these definitions we can formulate the backwards correspondence to Theorem 1.4.2 as in

Theorem 1.4.5. For every frame $\mathcal{F}$ that is not reflexive, transitive, Euclidian, serial or symmetric, we can find a Kripke structure based on $\mathcal{F}$ which falsifies an instance of the corresponding axiom of Theorem 1.4.2.

Proof.

1. $\mathcal{F} = (W, E_a)$ being not reflexive implies that there is some $v \in W$ such that $(v, v) \notin E_a$. Set $P := W \setminus \{v\}$. Then, $\mathcal{K} = (\mathcal{F}, P), v \models K_a P \wedge \neg P$, so $\mathcal{K}, v \not\models K_a P \to P$.

2. $\mathcal{F} = (W, E_a)$ being not transitive implies that there are v, w, z with $(v, w), (w, z) \in E_a$ but $(v, z) \notin E_a$. With $P := W \setminus \{z\}$, we have that $\mathcal{K} = (\mathcal{F}, P), v \not\models K_a P \to K_a K_a P$.

3. $\mathcal{F} = (W, E_a)$ being not Euclidian means that there exist v, w, z with $(v, w), (v, z) \in E_a$ but $(w, z) \notin E_a$. Thus, for $P := W \setminus \{z\}$ it is $\mathcal{K} = (\mathcal{F}, P), v \not\models \neg K_a P \to K_a \neg K_a P$.

4. $\mathcal{F} = \{W, E_a)$ being not serial implies that there is some v with $vE_a = \emptyset$ and thus $\mathcal{K}, v \not\models \neg K_a 0$ for any $\mathcal{K}$ based on $\mathcal{F}$.

5. $\mathcal{F} = \{W, E_a)$ being not symmetric means that there are v, w with $(v, w) \in E_a$ but $(w, v) \notin E_a$. For $P := \{v\}$ we have that $(\mathcal{F}, P), v \models P \wedge \neg K_a \neg K_a \neg P$.

□

1.5 Completeness

So far, we attempted to characterize the properties of knowledge, that is, the properties of Kripke structures modeling knowledge of actors in a certain setting, in terms of valid formulae. All we did though was to list *some* valid properties. There is little reason for us to believe that there are no additional properties (that are not consequences of those we listed). In this section, it is our goal to give a characterization of the valid properties of knowledge which is actually complete. In other words, we want to find a *sound* and *complete axiomatization* of Kripke structures that speak about knowledge. Remember that an axiomatization or *axiom system* is a collection of axioms and inference rules. A *proof* in the axiom system consists of a sequence of formulae, each of which is either an instance of an axiom in AX or follows by an application of an inference rule. A proof is said to be a *proof of the formula* φ if the last formula in the proof is φ. We say φ is *provable* in AX, and write $AX \vdash \varphi$, if there is a proof of φ in AX. Now, an axiomatization is *complete* with respect to a class of structures if every formula that is valid in this class of structures is provable in AX. It is called sound if every formula which is provable in AX is valid with respect to the class of structures.

We start by defining several axiom systems. The first one we call *(K)* and it consists of the following two axioms and two inference rules for a fixed set of agents A:

- All instances of propositional tautologies *(PL)*.
- $K_a\psi \wedge K_a(\psi \to \varphi) \to K_a\varphi$ (K).
- From ψ and $\psi \to \varphi$ infer φ (R1) (Modus Ponens).
- From φ infer $K_a\varphi$ (R2) (Knowledge Generalization).

We can add further axioms and obtain some more axiom systems:

- Add (T) (Knowledge Axiom) to (K) and call it *(T)*.
- Add (D) (Consistency Axiom) to (T) and call it *(D)*.
- Add (4) (Positive Introspection Axiom) to (D) and call it *(S4)*.

- Add (5) (Negative Introspection Axiom) to (S4) and call it *(S5)*.

The names of these axiom systems denote its most important axiom. We will see that each of them is a sound and complete axiomatization of a certain class of Kripke structures.

First, we intend to show that (K) is a complete axiomatization for modal logic in general, i. e., that if $\mathcal{K} \models \varphi$ for all Kripke structures $\mathcal{K}$ (with E_a not necessarily an equivalence relation), then φ can be derived by (K). To this end we want to introduce some notation. An axiomatization for a logic $\mathcal{L}$ we denote by AX. We write $AX \vdash \varphi$ if AX proves φ. We say that φ is *AX–consistent* if $AX \not\vdash \neg\varphi$, and that $\{\varphi_1, \ldots, \varphi_n\}$ is AX–consistent if $\varphi_1 \wedge \ldots \wedge \varphi_n$ is. Further, $\Phi \subseteq \mathcal{L}$ is AX–consistent if all finite subsets $\Phi_0 \subseteq \Phi$ are. Φ is a *maximal AX–consistent* set in $\mathcal{L}$ if Φ is AX–consistent but for every $\varphi \in \mathcal{L} \setminus \Phi$, the set $\Phi \cup \{\varphi\}$ is not AX–consistent.

With this we show the following rather technical lemma.

Lemma 1.5.1. For any axiomatization AX that includes every instance of (PL) and (R1), for a countable logic $\mathcal{L}$ that is closed under propositional correctives (so that if φ and ψ are in $\mathcal{L}$, then so are $\varphi \wedge \psi$ and $\neg\varphi$), every AX–consistent set Φ_0 can be extended to a maximal AX–consistent set Φ. Further, maximal AX–consistent sets Φ have the following properties:

1. If $\varphi \in \mathcal{L}$, then either $\varphi \in \Phi$ or $\neg\varphi \in \Phi$,
2. $\varphi \wedge \psi \in \Phi \Leftrightarrow \varphi, \psi \in \Phi$,
3. if $\psi \in \Phi$ and $\psi \to \varphi \in \Phi$, then also $\varphi \in \Phi$,
4. if $AX \vdash \varphi$, then $\varphi \in \Phi$.

Proof. Fix $\psi_1, \psi_2, \ldots$, an enumeration of $\mathcal{L}$. Construct a sequence $\Phi_0 \subseteq \Phi_1 \subseteq \cdots$ of AX–consistent sets, where Φ_0 is the set we want to extend to a maximal AX–consistent set, and $\Phi_{i+1} := \Phi_i \cup \{\psi_i\}$ if this set is AX–consistent, and $\Phi_{i+1} := \Phi_i$ otherwise. Now, let $\Phi := \bigcup_{i \in \mathbb{N}} \Phi_i$. Because all its finite subsets are contained in some AX–consistent Φ_i and are thus AX–consistent, Φ is AX–consistent. On the other hand, if $\psi \notin \Phi$, then $\psi = \psi_i$ for some i and $\Phi_i \cup \{\psi\}$ was not AX–consistent. Hence, $\Phi \cup \{\psi_i\}$ is not AX–consistent. It follows that Φ is a maximal AX–consistent set.

To see that maximal AX–consistent sets have all the properties we claimed, let Φ be an arbitrary maximal AX–consistent set.

1. As Φ is consistent, φ and $\neg\varphi$ cannot be both its member. If none is in Φ, then because of its maximality neither $\Phi \cup \{\varphi\}$ nor $\Phi \cup \{\neg\varphi\}$ are AX–consistent. By definition of AX–consistency of sets this means that

there exist $\psi_1, \ldots, \psi_k, \psi'_1, \ldots, \psi'_\ell \in \Phi$ such that $AX \vdash \neg(\psi_1 \wedge \ldots \wedge \psi_k \wedge \varphi)$ and $AX \vdash \neg(\psi'_1 \wedge \ldots \wedge \psi'_\ell \wedge \neg\varphi)$. With purely propositional reasoning we can conclude that $AX \vdash \neg((\psi_1 \wedge \ldots \wedge \psi_k \wedge \varphi) \vee (\psi'_1 \wedge \ldots \wedge \psi'_\ell \wedge \neg\varphi))$ and thus that $\Phi \cup \{\varphi \vee \neg\varphi\}$ is AX–inconsistent, implying that Φ is AX–inconsistent, which is contradicts our assumption.

2. If $\varphi \wedge \psi \in \Phi$, then $\varphi \in \Phi$ because otherwise $\neg\varphi \in \Phi$ and Φ would not be AX–consistent. Conversely, if $\varphi, \psi \in \Phi$, then $\varphi \wedge \psi \in \Phi$ because otherwise $\neg(\varphi \wedge \psi) \in \Phi$ and $\{\varphi, \psi, \neg(\varphi \wedge \psi)\} \subseteq \Phi$ is AX–inconsistent.

Like 2., 3. and 4. can be shown easily using 1. □

We continue with proving the actual Theorem.

Theorem 1.5.2. (K) is a sound and complete axiomatization for modal logic.

Proof. Given that the axioms of (K) are valid and its inference rules are sound, it is straightforward to prove that (K) is sound by induction on the length of the proof

To prove completeness, we must show that every modal formula which is valid with respect to every suitable $\mathcal{K}$ is provable in (K). It suffices to prove *model existence* for (K)–consistent formulae, i. e.:

For every (K)–consistent $\varphi \in \mathrm{ML}$, we find $\mathcal{K}$ and $w \in \mathcal{K}$ such that $\mathcal{K}, w \models \varphi$.

Suppose we can prove this and φ is a valid modal formula. If φ is not provable in (K), then neither is $\neg\neg\varphi$, so, by definition, $\neg\varphi$ is (K)–consistent. It follows with our assumption that $\neg\varphi$ is satisfiable, contradicting the validity of φ.

We prove model existence for (K)–consistent formulae by constructing a generic model $\mathcal{K}_c$ called the *canonical structure.* It has a node w_Φ for every maximal (K)–consistent set Φ:

$$W_c := \{w_\Phi : \Phi \text{ is maximal (K)–consistent in ML}\}.$$

To define the edges, for $\Phi \subseteq \mathrm{ML}$ we let $\Phi/a := \{\varphi : K_a\varphi \in \Phi\}$ be the set of all formulae in Φ which agent a knows. Now,

$$E_a := \{(w_\Phi, w_\Psi) \in W_c \times W_c \mid \Phi/a \subseteq \Psi\},$$

$$P_i := \{w_\Phi \in W_c \mid P_i \in \Phi\}.$$

We claim that for all Φ and Ψ, we have that

$$\mathcal{K}_c, w_\Phi \models \varphi \Leftrightarrow \varphi \in \Phi.$$

To show this, we proceed by induction on the structure of the formulae.

- If $\varphi = P_i$ is an atom, the claim is immediate from the definition of $\mathcal{K}_c$.
- If $\varphi = \neg\psi$, then $\mathcal{K}_c, w_\Phi \models \varphi \Leftrightarrow \mathcal{K}_c, w_\Phi \not\models \psi \overset{\text{I.H.}}{\Longleftrightarrow} \psi \notin \Phi \overset{\Phi \text{ max. cons.}}{\Longleftrightarrow} \varphi \in \Phi$.
- If $\varphi = \psi \wedge \vartheta$, then $\mathcal{K}_c, w_\Phi \models \varphi \Leftrightarrow \mathcal{K}_c, w_\Phi \models \psi$ and $\mathcal{K}_c, w_\Phi \models \vartheta \overset{\text{I.H.}}{\Longleftrightarrow} \psi, \vartheta \in \Phi \overset{\Phi \text{ max. cons.}}{\Longleftrightarrow} \varphi \in \Phi$.
- For the interesting case, assume that φ is of the form $K_a\psi$.

  "$\Leftarrow$": If $\varphi \in \Phi$, then $\psi \in \Phi/a$, and, by definition of E_a, we have that $\psi \in \Psi$ for all $w_\Psi \in w_\Phi E_a$. Thus, by the induction hypothesis, $\mathcal{K}_c, w_\Psi \models \psi$ for all $w_\Psi \in w_\Phi E_a$. This directly implies $\mathcal{K}_c, w_\Phi \models \varphi$.

  "$\Rightarrow$": Assume $\mathcal{K}_c, w_\Phi \models \varphi$. It follows that the set $\Phi/a \cup \{\neg\psi\}$ is not (K)–consistent. Otherwise, by Lemma 1.5.1, it would have a maximal (K)–consistent extension Ψ and by construction $w_\Psi \in w_\Phi E_a$. With the induction hypothesis, $\mathcal{K}_c, w_\Psi \models \neg\psi$. Hence $\mathcal{K}_c, w_\Phi \not\models K_a\psi = \varphi$, which is a contradiction to our original assumption. Since $\Phi/a \cup \{\neg\psi\}$ is not (K)–consistent, there must be a finite subset $\{\varphi_1, \ldots, \varphi_k, \neg\psi\}$ which is not (K)–consistent. With propositional reasoning , we have

  $$(\mathrm{K}) \vdash \varphi_k \to (\varphi_{k-1} \to (\ldots \to (\varphi_1 \to \psi) \ldots)).$$

  By Knowledge Generalization, we have

  $$(\mathrm{K}) \vdash K_a(\varphi_k \to (\varphi_{k-1} \to (\ldots \to (\varphi_1 \to \psi) \ldots))).$$

  To continue, we need the following

  Lemma 1.5.3.

  $$\begin{aligned}(\mathrm{K}) \vdash\ & K_a(\varphi_k \to (\varphi_{k-1} \to (\ldots \to (\varphi_1 \to \psi) \ldots))) \to \\ & K_a\varphi_k \to (K_a\varphi_{k-1} \to (\ldots \to (K_a\varphi_1 \to K_a\psi) \ldots)).\end{aligned}$$

  Proof. Set

  $$\alpha_k := \varphi_k \to (\varphi_{k-1} \to (\ldots \to (\varphi_1 \to \psi) \ldots)) = \varphi_k \to \alpha_{k-1},$$

  and

  $$\begin{aligned}\beta_k :=\ & K_a\varphi_k \to (K_a\varphi_{k-1} \to (\ldots \to (K_a\varphi_1 \to K_a\psi) \ldots)) \\ =\ & K_a\varphi_k \to \beta_{k-1}.\end{aligned}$$

With this we can rewrite the claim of the lemma to

$$(\mathrm{K}) \vdash K_a\alpha_k \to \beta_k.$$

Now, if $k = 1$, the claim is $(\mathrm{K}) \vdash K_a(\varphi_1 \to \psi) \to (K_a\varphi_1 \to K_a\psi)$, which is up to propositional equivalence just an instance of the Distribution Axiom $K_a(\varphi \to \psi) \land K_a\varphi \to K_a\psi$.

If $k > 1$, then by the induction hypothesis we know that $(\mathrm{K}) \vdash K_a\alpha_{k-1} \to \beta_{k-1}$. Similarly like for $k = 1$, we use the version of the Distribution Axiom $(\mathrm{K}) \vdash K_a(\varphi_k \to \alpha_{k-1}) \to (K_a\varphi_k \to K_a\alpha_{k-1})$. When replacing $K_a\alpha_{k-1}$ by β_{k-1}, we get $(\mathrm{K}) \vdash K_a(\varphi_k \to \alpha_{k-1}) \to (K_a\varphi_k \to \beta_{k-1})$. □

With Modus Ponens, the claim from the lemma together with the claim right before it can be inferred to

$$K_a\varphi_k \to (K_a\varphi_{k-1} \to (\ldots \to (K_a\varphi_1 \to K_a\psi)\ldots)).$$

Since $\varphi_1, \ldots, \varphi_k \in \Phi/a$, it follows that $K_a\varphi_1, \ldots, K_a\varphi_k \in \Phi$. By part 4. of Lemma 1.5.1 applied repeatedly, we have $\varphi = K_a\psi \in \Phi$.

□

Note that the canonical structure $\mathcal{K}_c$ is infinte. In the following we will see that for a given formula φ we can refine the construction so as to produce a finite model for this formula. The main idea is to only use sets of subformulae or its negations of φ as nodes in the model. Towards this end, we define

$$Sf(\varphi) := \{\psi : \psi \text{ is a subformula of } \varphi\},$$

and

$$Sf^{\neg}(\varphi) := Sf(\varphi) \cup \{\neg\psi : \psi \in Sf(\varphi)\}.$$

Further,

$$\mathit{Con}(\varphi) := \{\Phi \subseteq Sf^{\neg}(\varphi) \mid \Phi \text{ is maximal (K)–consistent inside } Sf^{\neg}(\varphi)\}.$$

A proof almost identical to that of Lemma 1.5.1 can be used to show that

- every (K)–consistent subset of $Sf^{\neg}(\varphi)$ can be extended to a set $\Phi \in \mathit{Con}(\varphi)$,
- and every $\Phi \in \mathit{Con}(\varphi)$ contains either φ or $\neg\varphi$ for every $\varphi \in Sf^{\neg}(\varphi)$.

Now, the construction of the model for φ is similar to $\mathcal{K}_c$. If we call it $\mathcal{K}_\varphi$, then $\mathcal{K}_\varphi := (W_\varphi, (E_a)_{a \in A}, (P_i)_{i \in I})$, where $W_\varphi := \{w_\Phi : \Phi \in Con(\varphi)\}$, $E_a := \{(w_\Phi, w_\Psi) : \Phi/a \subseteq \Psi\}$, and $P_i := \{w_\Phi : P_i \in \Phi\}$. Analogously, we can then prove that $K_\varphi, w_\Phi \models \varphi \Leftrightarrow \varphi \in \Phi$ for all $\varphi \in Sf^\neg(\varphi)$.

Corollary 1.5.4. If $\varphi \in$ ML is (K)–consistent, then φ is satisfiable in a Kripke structure with at most $2^{|Sf(\varphi)|} \leq 2^{|\varphi|}$ states. We call this *small model property.*

Analogous statements are true for (T)–, (S4)–, and (S5)–consistent formulae. For (T)–consistent formulae φ, consider $\mathcal{K}_\varphi$ whose nodes are w_Φ for the maximal (T)–consistent subsets Φ of $Sf^\neg(\varphi)$. Since we have in (T) the axiom $K_a\varphi \to \varphi$ it follows that $\Phi/a \subseteq \Phi$. So E_a is reflexive. Hence, every (T)–consistent φ is satisfiable in a *reflexive* Kripke structure with at most $2^{|\varphi|}$ states.

For an (S4)–consistent φ we need a more involved construction to guarantee the *transitivity* of E_a. Let us first illustrate the difficulty here. Consider $\varphi := K_aP$ and the following maximal (S4)–consistent subsets of $Sf^\neg(\varphi)$: $\Phi_1 := \{K_aP, P\}, \Phi_2 := \{\neg K_aP, P\}, \Phi_3 := \{\neg K_aP, \neg P\}$. We have $\Phi_1/a \subseteq \Phi_2$ and $\Phi_2/a \subseteq \Phi_3$ but $\Phi_1/a \not\subseteq \Phi_3$, so if we used the same construction again, $\mathcal{K}_\varphi$ would not be transitive. To solve this, take $E_a := \{(w_\Phi, w_\Psi) : \Phi/a \subseteq \Psi/a\}$. Then, $\Phi/a \subseteq \Psi/a \subseteq \Theta/a \Rightarrow \Phi/a \subseteq \Theta/a$. Thus, every (S4)–consistent φ is satisfiable in a $\mathcal{K}$ of size at most $2^{|\varphi|}$ which is reflexive and transitive.

For (S5), $E_a := \{(w_\Phi, w_\Psi) : \Phi/a = \Psi/a\}$ is an appropriate equivalence relation.

Corollary 1.5.5. For modal logic:

1. (T) is a sound and complete axiomatization with respect to reflexive Kripke structures.

2. (S4) is a sound and complete axiomatization with respect to reflexive and transitive Kripke structures.

3. (S5) is a sound and complete axiomatization with respect to reflexive, symmetric and transitive Kripke structures.

Proof. Completeness follows from the respective small model property. Here we use again that for showing completeness it is enough to show model existence for formulae consistent with the respective axiom system (as explained at the beginning of the proof of Theorem 1.5.2). To see soundness, remember Theorem 1.4.2.

We would like to add that the axiom $\varphi \to K_a\neg K_a\neg\varphi$ is a consequence of (D) and (5). This refers to the correspondence of this axiom to the symmetry of E_a like shown in Theorem 1.4.2 and Theorem 1.4.5. In that sense we uphold that (T) corresponds to reflexivity, (4) to transitivity, and $\varphi \rightarrow K_a \neg K_a \neg \varphi$ to symmetry. □

Finally, without proof, we state the following theorem.

Theorem 1.5.6. Denote by $(\mathrm{K})^{\mathrm{C}}$, $(\mathrm{T})^{\mathrm{C}}$, $(\mathrm{S4})^{\mathrm{C}}$, and $(\mathrm{S5})^{\mathrm{C}}$ the respective axiom system extended by

- $E_G\varphi \leftrightarrow \bigwedge_{i \in G} K_i\varphi$ (C1),
- $C_G\varphi \rightarrow E_G(\varphi \wedge C_G\varphi)$ (C2),
- From $\varphi \rightarrow E_G(\psi \wedge \varphi)$ infer $\varphi \rightarrow C_g\psi$ (Induction Rule).

Then, for modal logic extended by the common knowledge operator:

1. $(\mathrm{K})^{\mathrm{C}}$ is a sound and complete axiomatization (with respect to arbitrary Kripke structures).
2. $(\mathrm{T})^{\mathrm{C}}$ is a sound and complete axiomatization with respect to reflexive Kripke structures.
3. $(\mathrm{S4})^{\mathrm{C}}$ is a sound and complete axiomatization with respect to reflexive and transitive Kripke structures.
4. $(\mathrm{S5})^{\mathrm{C}}$ is a sound and complete axiomatization with respect to reflexive, symmetric and transitive Kripke structures.

1.6 Complexity

The satisfiability problem of an axiom system AX is the question whether a given formula is satisfiable within the class of structures axiomatized by AX, or equivalently, whether the formula is AX–consistent. From the small model property it follows that the satisfiability problems for all considered variants of modal logic are decidable in NEXPTIME. Indeed, it is not difficult to see that their evaluation problem (given $\mathcal{K}, w$ and φ: decide whether $\mathcal{K}, w \models \varphi$) is decidable in time $\mathcal{O}(||\mathcal{K}|| \cdot |\varphi|)$.

This upper bound is probably not optimal. As a matter of fact, the following complexities hold:

NP–complete:	(S5) where $\lvert A\rvert = 1$
PSPACE–complete:	(K),(T),(S4) with arbitrary $\lvert A\rvert$ (S5) with $\lvert A\rvert > 1$
EXPTIME–complete:	$(\mathrm{K})^{\mathrm{C}}$, $(\mathrm{T})^{\mathrm{C}}$ with arbitrary A $(\mathrm{S4})^{\mathrm{C}}$, $(\mathrm{S5})^{\mathrm{C}}$ with $\lvert A\rvert > 1$

Figure 1.3: Complexities of satisfiability problems of various axiom systems.

We are going to provide a proof for SAT(S5) being NP–complete if $|A| = 1$, and SAT(K) being PSPACE–complete.

Theorem 1.6.1. SAT(S5) for $|A| = 1$ is NP–complete.

Proof. NP–hardness holds since clearly SAT(S5) must be at least as hard as the satisfiability problem for propositional logic: a propositional formula is satisfiable if, and only if, it is satisfiable as a formula in modal logic with propositional variables translated to atomic propositions by a Kripke structure with whatever restrictions on its edge relations.

To prove that SAT(S5) $\in$ NP for $|A| = 1$, we observe that for a Kripke structure $\mathcal{K}$ whose (single) edge relation E_a is an equivalence relation like it is the case for those axiomatized by (S5), if we have that $\mathcal{K}, v \models \varphi$, then also $\mathcal{K} \upharpoonright_{[w]}, w \models \varphi$, where $[w]$ is the equivalence class of w with respect to E_a, and $\mathcal{K} \upharpoonright_{[w]}$ is the restriction of $\mathcal{K}$ to $[w]$. This is because every node which is relevant for the evaluation of φ at v in $\mathcal{K}$ is reachable from v and thus, since $|A| = 1$ and E_a is an equivalence relation, is in the same equivalence class as v. In fact, every component of $\mathcal{K}$ is an E_a–clique.

Now, assume that a formula φ is satisfiable at node w in a Kripke structure $\mathcal{K} = (W, E_a, (P_i)_{i\in I})$, where $E_a = W \times W$. Given such a model, let

$$\Phi := \{K_a\psi \in Sf(\varphi) \mid \mathcal{K}, w \models \neg K_a\psi\}.$$

For every $K_a\psi \in \Phi$, there is a node $v_\psi \in W$ with $\mathcal{K}, v_\psi \models \neg\psi$. Let $\mathcal{K}'$ be the restriction of $\mathcal{K}$ to $\{w\} \cup \{v_\psi : K_a\psi \in \Phi\}$. We claim that $\mathcal{K}', w \models \varphi$. Since $|\mathcal{K}'| \leq |\varphi|$, this would prove the theorem.

Concerning the claim, we show via induction that for all $\psi \in Sf(\varphi)$ and all $v \in W'$ (in particular φ and w themselves) we have that

$$\mathcal{K}, v \models \psi \Leftrightarrow \mathcal{K}', v \models \psi.$$

This induction is trivial in all cases except for $\psi = K_a\vartheta$:

"$\Rightarrow$" $\mathcal{K}, v \models K_a\vartheta \overset{\text{Def.}}{\Longleftrightarrow} \mathcal{K}, z \models \vartheta$ for all $z \in W \overset{W' \subseteq W}{\Longrightarrow} \mathcal{K}, z \models \vartheta$ for all $z \in W' \overset{\text{I.H.}}{\Longleftrightarrow} \mathcal{K}', z \models \vartheta$ for all $z \in W' \overset{\text{Def.}}{\Longrightarrow} \mathcal{K}', v \models K_a\vartheta$.

"$\Leftarrow$" $\mathcal{K}, v \not\models K_a\vartheta \overset{\text{Def.}}{\Longrightarrow} \mathcal{K}, v \models \neg K_a\vartheta \overset{\text{Def.}}{\Longrightarrow} \mathcal{K}, z \models \neg\vartheta$ for some $z \overset{(w,z)\in E_a}{\Longrightarrow} \mathcal{K}, w \models \neg K_a\vartheta \overset{\text{Def.}}{\Longrightarrow} K_a\vartheta \in \Phi \overset{\text{Def.}}{\Longrightarrow} \exists v_\psi \in \mathcal{K}' : \mathcal{K}, v_\psi \models \neg\vartheta \overset{\text{I.H.}}{\Longrightarrow} \mathcal{K}', v_\psi \models \neg\vartheta \overset{(v,v_\psi)\in E_a}{\Longrightarrow} \mathcal{K}', v \models \neg K_a\vartheta$.

□

Theorem 1.6.2. SAT(K)=SAT(ML) can be decided in PSPACE.

Proof. Let φ be a modal formula, for technical reasons in positive normal form, i. e., negations occur only in front of atoms. In order to not lose expressive power, we now have to allow also the $\langle a\rangle$ operator.

First, we define the *closure* $Cl^*(\varphi)$ of a formula φ, by

- $(\varphi, 0) \in Cl^*(\varphi)$,
- $(\psi \circ \vartheta, i) \in Cl^*(\varphi) \Rightarrow (\psi, i), (\vartheta, i) \in Cl^*(\varphi)$, for $\circ = \wedge, \vee$,
- $(\langle a\rangle\psi, i) \in Cl^*(\varphi) \Rightarrow (\psi, i+1) \in Cl^*(\varphi)$,
- $([a]\psi, i) \in Cl^*(\varphi) \Rightarrow (\psi, i+1) \in Cl^*(\varphi)$,
- $(\neg P_j, i) \in Cl^*(\varphi) \Rightarrow (P_j, i) \in Cl^*(\varphi)$.

Based on that, for each $i = 0, \ldots, md(\varphi) := \arg\max_i \exists\psi(\psi, i) \in CL^*(\varphi)$, we define

$$Cl^{(i)} := \{\psi : (\psi, i) \in Cl^*(\varphi)\} \subseteq \text{ML}.$$

We call a $\Gamma \subseteq Cl^{(i)}(\varphi)$ *propositionally correct* if for all

- literals $(\neg)P_j \in Cl^{(i)}(\varphi)$, we have $P_j \in \Gamma \Leftrightarrow \neg P_j \notin \Gamma$,
- conjunctions $(\psi \wedge \vartheta) \in \Gamma \Rightarrow \psi, \vartheta \in \Gamma$,
- disjunctions $(\psi \vee \vartheta) \in \Gamma \Rightarrow \psi \in \Gamma$ or $\vartheta \in \Gamma$.

It is easy to see that we can check efficiently whether a given $\Gamma \subseteq Cl^{(i)}(\varphi)$ is propositionally correct. With this we formulate the following algorithm "Check(Γ, i)".

```
Known: $\varphi$ and $Cl^{(j)}(\varphi)$ for $j = 0, \ldots, md(\varphi)$
Given: $\Gamma \subseteq Cl^{(i)}(\varphi)$
Check whether $\Gamma$ is propositionally correct. If not, reject;
for $\langle a \rangle \psi \in \Gamma$ do
    Guess $\Gamma_\psi \subseteq Cl^{(i+1)}(\varphi)$ with
        • $\psi \in \Gamma_\psi$,
        • $\{\vartheta : [a]\vartheta \in \Gamma\} \subseteq \Gamma_\psi$;
    Call Check($\Gamma_\psi, i+1$). If it rejects, reject;
end
Accept;
```

This is a non–deterministic algorithm requiring space $\mathcal{O}(|\varphi|^2)$. We claim that we can decide SAT(K) with this algorithm. The procedure is as follows: For a given $\varphi \in$ ML,

1. compute $Cl^*(\varphi)$,
2. guess $\Gamma_\varphi \subseteq Cl^0(\varphi)$ with $\varphi \in \Gamma_\varphi$,
3. check $(\Gamma_\varphi, 0)$.

Let us prove this:

"⇒": Assume φ is satisfiable with $\mathcal{K}, u \models \varphi$. Then, the procedure accepts with the following guesses:

The first guess is $\Gamma_\varphi := \{\psi \in Cl^{(0)}(\varphi) \mid \mathcal{K}, u \models \psi\}$. This is obviously a subset of $Cl^{(0)}(\varphi)$ and $\varphi \in Cl^{(0)}(\varphi)$ is a member. Also, by definition, it is propositionally correct.

We continue such that for every call of Check(Γ, i) some node $v(\Gamma)$ is fixed. For Check$(\Gamma_\varphi, 0)$ set $v(\Gamma_\varphi) := u$. Inside Check(Γ, i), for all $\langle a \rangle \psi \in \Gamma$ select some $w(\Gamma_\psi)$ with $(v(\Gamma), w(\Gamma_\psi)) \in E_a$ and $\mathcal{K}, w(\Gamma_\psi) \models \psi$ (which is possible since $\mathcal{K}, v(\Gamma) \models \langle a \rangle \psi$), and let $\Gamma_\psi := \{\vartheta : \mathcal{K}, w(\Gamma_\psi) \models \vartheta\}$. This is a valid guess since the set is propositionally correct, and it contains ψ and all ϑ with $[a]\vartheta \in \Gamma_\psi$. The latter because from $[a]\vartheta \in \Gamma$ it follows $\mathcal{K}, v(\Gamma) \models [a]\vartheta$ which implies that $\mathcal{K}, w(\Gamma_\psi) \models \vartheta$.

"⇐": Now, assume that the algorithm accepts. Based on this, we construct a finite tree model of φ. To this end, for every call of Check(Γ, i) we create a node v and set $\Gamma_v := \Gamma$. In the first call Check$(\Gamma_\varphi, 0)$, we create the root u with $\Gamma_u := \Gamma_\varphi$. Further, inside Check$(\Gamma, i)$, if the current node is v (with $\Gamma_v = \Gamma$), for each $\langle a\rangle\psi \in \Gamma$ create a new a–successor $w(\psi)$ of v and set $\Gamma_{w(\psi)} := \Gamma_\psi$.

This produces a finite tree $(V, (E_a)_{a\in A})$. Set $P'_j := \{v : P_j \in \Gamma_v\}$.

We claim that $\mathcal{T} := (V, (E_a)_{a\in A}, (P'_j)_{j\in I}), u \models \varphi$. By induction over the structure of the formula we will show that

$$\psi \in \Gamma_v \Rightarrow \mathcal{T}, v \models \psi.$$

In particular, since the algorithm accepts and thus $\varphi \in \Gamma_\varphi = \Gamma_u$, this implies $\mathcal{T}, u \models \varphi$. We know that $\psi \in \Gamma_v \subseteq Cl^{(i)}(\varphi)$ is a positive Boolean combination of $(\neg)P_j$ and $\langle a\rangle\vartheta, [a]\vartheta$ with $\vartheta \in Cl^{(i+1)}$, hence these are the cases we have to consider.

(I.B.) If $\psi = P_j \in \Gamma_v$, we have $\mathcal{T}, v \models \psi$ by definition of P'_j. If $\psi = \neg P_j \in \Gamma_v$, we have $\mathcal{T}, v \models \psi$ by definition of P'_j and the fact that because Γ_v is propositionally correct $P_j \notin \Gamma_v$.

(I.S.) As Γ_v is propositionally correct, for $\psi = \vartheta \wedge \chi$ and $\psi = \vartheta \vee \chi$ we can use the induction hypothesis immediately. If $\psi = \langle a\rangle\vartheta$, then we created an a–successor $w(\vartheta) \in V$ of v and set $\Gamma_{w(\vartheta)} := \Gamma_\vartheta$. By definition of the algorithm we have that $\vartheta \in \Gamma_\vartheta = \Gamma_{w(\vartheta)}$ which, with the induction hypothesis, implies that $\mathcal{T}, w(\vartheta) \models \vartheta$, and hence $\mathcal{T}, v \models \psi$. Finally, consider the case of $\psi = [a]\vartheta$. Remember that we constructed for each $\langle a\rangle\chi \in \Gamma_v$ an a–successor $w(\chi)$ of v and set $\Gamma_{w(\chi)} := \Gamma_\chi$. These are all a–successors the node v has. Because of how the algorithm chose Γ_χ, we know that $\{\xi : [a]\xi \in \Gamma_v\} \subseteq \Gamma_\chi = \Gamma_{w(\chi)}$. In particular, $\vartheta \in \Gamma_w$ for all successors w of v, which, with the induction hypothesis, implies $\mathcal{T}, w \models \vartheta$ for all successors w of v. We conclude that $\mathcal{T}, v \models \psi$.

□

The following is a nice general consequence.

Corollary 1.6.3. Every satisfiable formula $\varphi \in$ ML has a tree model of height not larger than $md(\varphi)$ and branching degree not larger than $|\varphi|$.

Before we prove PSPACE–hardness of SAT(K), we show that there are modal formulae of ever increasing length which force its models to be exponential with respect to their length. The proof of this prepares for the proof regarding PSPACE–hardness.

Theorem 1.6.4. There is a sequence of formulae $(\varphi_n)_{n\in\mathbb{N}}$ in ML such that

1. $|\varphi_n| \in \mathcal{O}(n^2 \cdot \log n)$,

2. φ_n is satisfiable,

3. if $\mathcal{K}, w \models \varphi_n$, then $|\mathcal{K}| \geq 2^n$.

Proof. We use variables $X_1, \ldots, X_n, Y_0, \ldots, Y_{n+1}$ and define

$$\varphi_n := Y_0 \wedge \neg Y_1 \wedge \bigwedge_{i=0}^{n} [a]^i (\alpha \wedge \beta \wedge \gamma),$$

where

$$\alpha := \bigwedge_{i=1}^{n+1} Y_i \to Y_{i-1},$$

$$\beta := \bigwedge_{i=1}^{n} Y_i \to ((X_i \to [a]X_i) \wedge (\neg X_i \to [a]\neg X_i)),$$

$$\gamma := \bigwedge_{i=1}^{n-1} (Y_i \wedge \neg Y_{i+1}) \to (\langle a \rangle \gamma_i^+ \wedge \langle a \rangle \gamma_i^-),$$

$$\gamma_i^+ := Y_{i+1} \wedge \neg Y_{i+2} \wedge X_{i+1}, \text{ and } \gamma_i^- := Y_{i+1} \wedge \neg Y_{i+2} \wedge \neg X_{i+1}.$$

Note that $|\varphi_n| \in \mathcal{O}(n^2 \cdot \log n)$. The logarithmic term stems from the fact that we have to binary encode the counters of the variables X_i, Y_i. Now, the intended model of this formula is a full binary tree of heigt n, defined as $\mathcal{T}_n := (W, E_a, (X_i)_{1 \leq i \leq n}, (Y_i)_{0 \leq i \leq n+1})$, with $W := \{0,1\}^{\leq n}$, $E_a := \{(u,v) : v = u0 \text{ or } v = u1\}$, $Y_i = \{v : |v| \geq i\}$, $X_i = \{v : |v| \geq i \text{ and the } i\text{–th bit of } v \text{ is } 1\}$. Indeed, it is an easy exercise to check that $\mathcal{T}_n, \varepsilon \models \varphi_n$, where ε is the so–called *empty word* $\{0,1\}^0$.

The trickier thing to show is that every model of φ_n has size at least 2^n. Towards this, we prove the following claim: Let $\mathcal{K}, u \models \varphi_n$. Then, for all $j = 0, \ldots, n$ and $w = w_1 \ldots w_j \in \{0,1\}^j$ there is a state $v(w) \in V$ that is reachable from u in j steps with

$$\mathcal{K}, v(w) \models Y_j \wedge \neg Y_{j+1} \wedge \bigwedge_{i \leq j, w_i = 1} X_i \wedge \bigwedge_{i \leq j, w_i = 0} \neg X_i.$$

By definition of these formulae, all the states $v(w)$ would be different and thus this is enough to show that $|\mathcal{K}| \geq 2^n = |\{0,1\}^n|$. If $j = 0$, meaning that $w = \varepsilon$, we simply choose $v(w) := 0$. If $j > 0$, then $w = w'w_j$, where $w' \in \{0,1\}^{j-1}$ and $w_j \in \{0,1\}$. By the induction hypothesis, there is a $v(w')$, reachable from u in $j-1$ steps and $(*)$ $\mathcal{K}, v(w') \models Y_{j-1} \wedge \neg Y_j \wedge \bigwedge_{i<j, w'_i=1} X_i \wedge \bigwedge_{i<j, w'_i=0} \neg X_i$. Now, because $\mathcal{K}, u$ is a model of φ_n and $v(w')$ is reachable from u in not more than n steps, we know that $\mathcal{K}, v(w') \models \alpha \wedge \beta \wedge \gamma$. With

$(*)$ we have $\mathcal{K}, v(w') \models Y_{j-1} \wedge \neg Y_j \wedge \gamma$ in particular. By definition of γ this implies that there are successors z^+, z^- of $v(w')$ such that $\mathcal{K}, z^+ \models \gamma^+_{j-1}$ and $\mathcal{K}, z^- \models \gamma^-_{j-1}$. We define $v(w) := z^+$ if $w_j = 1$ and $v(w) := z^-$ if $w_j = 0$. Then, by definition of $\gamma^+_{j-1}/\gamma^-_{j-1}$ we have $\mathcal{K}, v(w) \models Y_j \wedge \neg Y_{j+1} \wedge X_j/\neg X_j$, depending on whether $w_j = 1$ or $w_j = 0$. With $\mathcal{K}, v(w') \models \beta$ we can conclude that $\mathcal{K}, v(w) \models X_i \Leftrightarrow \mathcal{K}, v(w') \models X_i$ for $j = 1, \ldots, j-1$, so in particular $\mathcal{K}, v(w) \models \bigwedge_{i \leq j-1, w_i=1} X_i \wedge \bigwedge_{i \leq j-1, w_i=0} \neg X_i$. Altogether this shows the claim. □

Theorem 1.6.5. SAT(K) is PSPACE–complete.

Proof. We have already shown in Theorem 1.6.2 that SAT(K) $\in$ PSPACE. To prove hardness, we reduce QBF to it. That is, given a quantified Boolean formula $\varphi = Q_1 X_1 \ldots Q_n X_n \psi(X_1, \ldots, X_n)$ with $Q_i \in \{\exists, \forall\}$ and ψ a propositional formula, construct a modal formula φ^* such that φ is satisfiable if, and only if, φ^* is true.

The construction we use is almost the same as that from the proof of the previous Theorem 1.6.4 except that not all possible valuations of $X_1, \ldots, X_n$ appear as states of $\mathcal{K}$ but only those that are necessary to make the formula true.

We define

$$\varphi^* := Y_0 \wedge \neg Y_1 \wedge \bigwedge_{i=0}^{n} [a]^i (\alpha \wedge \beta \wedge \gamma_\varphi) \wedge [a]^n \psi,$$

where

$$\gamma_\varphi := \bigwedge_{i:Q_i=\forall} Y_i \wedge \neg Y_{i+1} \rightarrow \langle a \rangle \gamma_i^+ \wedge \langle a \rangle \gamma_i^- \wedge \bigwedge_{i:Q_i=\exists} Y_i \wedge \neg Y_{i+1} \rightarrow \langle a \rangle \gamma_i^+ \vee \langle a \rangle \gamma_i^-,$$

and $\alpha, \beta, \gamma_i^+, \gamma_i^-$ like in the proof of Theorem 1.6.2. Obviously, φ^* is computable from φ in polynomial time.

Let us show that if φ is true, then φ^* is in SAT(K). To this end, let $\varphi_i(X_1, \ldots, X_i) := Q_{i+1} X_{i+1} \ldots Q_n X_n \psi$ for $i = 0, \ldots, n$. Now, define $\mathcal{K}_\varphi := (W, E_a, (X_i)_{1 \leq i \leq n}, (Y_i)_{0 \leq i \leq n+1})$ as the restriction of the full binary tree $\mathcal{T}_n$ from the previous proof to $\{w \in \{0,1\}^i : i \leq n, \varphi_i[w_1, \ldots, w_i] \text{ is true}\}$. Then we already have $\mathcal{K}_\varphi, \varepsilon \models \varphi^*$. To see this, first note that because φ is satisfiable, ε is actually a node in $\mathcal{K}_\varphi$. Next, consider an arbitrary node $w = w_1 \ldots w_i$ with $0 \leq i < n$. We want to show that it is a model of γ_φ. By definition of $\mathcal{K}_\varphi$ we know that $\varphi_i[w]$ is true. If $Q_{i+1} = \forall$, this means that $\varphi_{i+1}[w, 0]$ as well as $\varphi_{i+1}[w, 1]$ is true which implies by definition of $\mathcal{K}_\varphi$ that both $w0$ and $w1$ exist as successor nodes in $\mathcal{K}_\varphi$, and hence $\mathcal{K}_\varphi, w \models \langle a \rangle \gamma_i^+ \wedge$

$\langle a\rangle\gamma_i^-$. If $Q_{i+1} = \exists$, then we can argue analogously for $\mathcal{K}_\varphi, w \models \langle a\rangle\gamma_i^+ \vee \langle a\rangle\gamma_i^-$. Hence, $\mathcal{K}_\varphi, w \models \gamma_\varphi$. Finally, every node with distance n from ε is a word of length n that satisfies ψ by definition of W.

For the reverse direction assume that φ^* is in SAT(K). We have to show that φ is true. Let $\mathcal{K}, u \models \varphi^*$. For each v in $\mathcal{K}$ and $i \leq n$, define $\varphi_i^v := Q_{i+1}X_{i+1} \ldots Q_nX_n\psi[w_1, \ldots, w_i]$, where $w_i := 1$ if $\mathcal{K}, v \models X_i$ and $w_i := 0$ if $\mathcal{K}, v \models \neg X_i$. We claim that if a node v is reachable in i steps from u and $\mathcal{K}, v \models Y_i \wedge \neg Y_{i+1}$, then φ_i^v is true. A proof can be conducted via induction over i, starting from $i = n$. So let v be a node reachable from u in n steps. Then, since $\mathcal{K}, u \models \varphi^* \Rightarrow \mathcal{K}, u \models [a]^n\psi$, we have that $\mathcal{K}, v \models \psi$ which is by definition of w_i the same as saying that $\psi(w_1, \ldots, w_n) = \varphi_n^v$ is true. Consider now the case $i < n$. We know that $\mathcal{K}, u \models [a]^i(\alpha \wedge \beta \wedge \gamma_\varphi)$, hence $\mathcal{K}, v \models \alpha \wedge \beta \wedge \gamma_\varphi$. Since also $\mathcal{K}, v \models Y_i \wedge \neg Y_{i+1}$, we know that $\mathcal{K}, v \models \langle a\rangle\gamma_i^+ \wedge \langle a\rangle\gamma_i^+$ if $Q_{i+1} = \forall$ and $\mathcal{K}, v \models \langle a\rangle\gamma_i^+ \vee \langle a\rangle\gamma_i^-$ if $Q_{i+1} = \exists$. $Q_{i+1} = \forall$: there exist v_0, v_1 such that $(v, v_0), (v, v_1) \in E_a$ and $\mathcal{K}, v_0 \models Y_{i+1} \wedge \neg Y_{i+2} \wedge \neg X_{i+1}$ and $\mathcal{K}, v_1 \models Y_{i+1} \wedge \neg Y_{i+2} \wedge X_{i+1}$. Together this implies with the induction hypothesis that $\varphi_{i+1}^{v0} = \varphi_{i+1}^v(X_{i+1}/0)$ and $\varphi_{i+1}^{v1} = \varphi_{i+1}^v(X_{i+1}/1)$. As a consequence, $\forall X_{i+1}\varphi_{i+1}^v = \varphi_i^v$ is true. The case for $Q_{i+1} = \exists$ can be handled analogously. Now, if $i = 0$, then $v = u$ and $\mathcal{K}, u \models Y_0 \wedge \neg Y_1$. Thus, $\mathcal{K}, u \models \varphi_0^v = \psi$, which shows the claim. □

Chapter 2

Reasoning about Uncertainty

Like with reasoning about knowledge, many representations of uncertainty start with a set of possible worlds. But now, in the case of uncertainty instead of knowledge, we also assign probabilities to possible worlds or sets of them, describing the assumed likelihood of that world, respectively one in that set of worlds, being the actual world. Capturing uncertainty like this with probabilities is perhaps the best–known approach and we begin this chapter by elaborating on it.

Example 2.0.6. Rolling a dice has six possible outcomes, so we describe this situation by six different worlds. Assuming a traditional dice is used, we assign to each world the probability 1/6. The probability of a set of worlds is just their probabilities added up.

For technical reasons, it is typically assumed that the set of sets of worlds to which probabilities are assigned ($\mathcal{F}$ in the following definition) satisfies some closure properties:

Definition 2.0.7. An *algebra* over W is a set $\mathcal{F} \subseteq \mathcal{P}(W)$ such that

1. $W \in \mathcal{F}$,
2. $U, V \in \mathcal{F} \Rightarrow \overline{U}, U \cup V \in \mathcal{F}$.

We call $\mathcal{F}$ a σ*–algebra* if it is closed under countable union.

In the current context, W would be the set of all possible worlds. Note that $\mathcal{F}$ can be a real subset of $\mathcal{P}(W)$, which means that a probability is assigned not necessarily to every set of worlds.

We also have some requirements concerning how probabilites are assigned to possible worlds (we call the function that maps worlds to their probability the *probability distribution* or *probability measure* and it is usually denoted by μ):

Definition 2.0.8. A *probability space* is a triple $(W, \mathcal{F}, \mu)$, where $\mathcal{F}$ is an algebra over W and $\mu\colon \mathcal{F} \to [0,1]$ is such that

1. $\mu(W) = 1$,
2. $U \cap V = \emptyset \Rightarrow \mu(U \cup V) = \mu(U) + \mu(V)$ for all $U, V \in \mathcal{F}$.

This definition directly implies $\mu(\emptyset) = 0$ since because W and $\emptyset$ are disjoint,

$$1 = \mu(W) = \mu(W \cup \emptyset) = \mu(W) + \mu(\emptyset) = 1 + \mu(\emptyset).$$

Further, 2. implies finite additivity: If $U_1, \ldots, U_n$ are pairwise disjoint, then

$$\mu(U_1 \cup \ldots \cup U_n) = \sum_{i=1}^{n} \mu(U_i).$$

With this we can conclude that for finite W and $\mathcal{F} = \mathcal{P}(W)$, any probabilty measure can be characterized as a function $\mu\colon W \to [0,1]$ such that $\sum_{w \in W} \mu(\{w\}) = 1$. That is, it suffices to define a probability measure μ only on the elements on W; it can then be uniquely extended to all subsets U of W by $\mu(U) := \sum_{w \in U} \mu(\{w\})$. If W is countably infinite, it is typically required that $\mathcal{F}$ is a σ–algebra, and that μ is countably additive, so that if $U_1, U_2, \ldots$ are pairwise disjoint sets in $\mathcal{F}$, then $\mu(\bigcup_{i \in \mathbb{N}} U_i) = \sum_{i \in \mathbb{N}} \mu(U_i)$.

2.1 Ramsey's Argument

One approach to justify the requirements of Definition 2.0.8 of a probability space is known under the name *Ramsey's Argument.* It is based on our intuition with respect to betting behaviour. Let us explain this argument in the following. Given a set W of worlds and a subset $U \subseteq W$, consider an agent who can evaluate bets of the form "If U happens (i. e. if the actual world is in U) then I win $1 - \alpha$ while if U does not happen I lose α", where $0 \leq \alpha \leq 1$. Denote such a bet by (U, α). The bet $(\overline{U}, 1 - \alpha)$ is called the *complementary* bet to (U, α); the agent loses $1 - \alpha$ if U happens and wins α if it does not.

As a convention, the larger a result is, the more desirable we consider it to be. With this, note that $(U, 0)$ is a "can't lose" bet for the agent. He wins 1 if U is the case and loses 0 otherwise. The bet becomes less and less attractive as α gets larger. Eventually, $(U, 1)$ is a "can't win" bet. The agent wins 0 if U is the case and loses 1 otherwise; the worst possible bet.

Now, suppose the agent must choose not just between individual bets but rather between sets of them. Instead of sets of bets we often say *books*. The *payoff* $||\mathcal{B}||_w$ of a book $\mathcal{B}$ on a world $w \in W$ is understood to be

$$\sum_{(U,\alpha)\in\mathcal{B}:w\in U}(1-\alpha) - \sum_{(U,\alpha)\in\mathcal{B}:w\notin U}\alpha.$$

We assume that the agent has a *preference order* $\succeq$ defined on sets of bets. Here, "preference" shall not mean "strictly better" but merely "at least as good". So $\mathcal{B} \succeq \mathcal{B}'$ means that the agent prefers $\mathcal{B}$ over $\mathcal{B}'$ or is indifferent between them. The preference order does not need to be total, that is, there might be books that are incomparable. However, if an agent is *rational* in the sense of the following four rationality postulates (R1) – (R4), then certainly some sets are comparable.

(R1) If $||\mathcal{B}||_w \geq ||\mathcal{B}'||_w$ for all $w \in W$ — we also say book $\mathcal{B}$ *guarantees to give at least as much as* $\mathcal{B}'$ — then $\mathcal{B} \succeq \mathcal{B}'$. If a book $\mathcal{B}$ *guarantees to give more than* $\mathcal{B}'$ (i. e., if $||\mathcal{B}||_w > ||\mathcal{B}'||_w$ for all $w \in W$), then $\mathcal{B} \succ \mathcal{B}'$.

Let us exemplify the conditions on α and β if $\mathcal{B}$ guarantees to give at least as much as $\mathcal{B}'$. For the sake of simplicity, we assume that the book $\mathcal{B}$ consists of only (U, α) and the book $\mathcal{B}'$ consists of only (V, β). Now, the following must hold:

1. If $U \cap V \neq \emptyset$, then $\alpha \leq \beta$. This is because on any $w \in U \cap V$, the agent wins $1 - \alpha = ||\mathcal{B}||_w$ with the bet (U, α), and $1 - \beta = ||\mathcal{B}'||_w$ with the bet (V, β). For $||\mathcal{B}||_w$ to be not less than $||\mathcal{B}'||_w$, this clearly requires $1 - \alpha \geq 1 - \beta$, implying $\alpha \leq \beta$.
2. If $\overline{U} \cap \overline{V} \neq \emptyset$, then $\alpha \leq \beta$ since for a $w \in \overline{U} \cap \overline{V}$ we must have $-\alpha = ||\mathcal{B}||_w \geq ||\mathcal{B}'||_w = -\beta$.
3. If $\overline{U} \cap V \neq \emptyset$, then $\alpha = 0$ and $\beta = 1$ since for a $w \in \overline{U} \cap V$ we must have $-\alpha = ||\mathcal{B}||_w \geq ||\mathcal{B}'||_w = 1 - \beta$.

(4.) If $U \cap \overline{V} \neq \emptyset$, then there is no condition on α, β since it always holds that $1 - \alpha = ||\mathcal{B}||_w \geq ||\mathcal{B}'||_w = -\beta$.

We note that for $(U, \alpha) \succeq (U, \beta)$ to hold, the above tells us that this is the case exactly if $\alpha \leq \beta$. This should seem reasonable.

(R2) Preferences are transitive: $\mathcal{B}_1 \succeq \mathcal{B}_2 \succeq \mathcal{B}_3 \Rightarrow \mathcal{B}_1 \succeq \mathcal{B}_3$.

(R3) One can always compare between complementary bets: for all (U, α), we have $(U, \alpha) \succeq (\overline{U}, 1 - \alpha)$ or $(U, \alpha) \preceq (\overline{U}, 1 - \alpha)$.

Note that this is indeed an assumption. In some cases it might also be considered reasonable to instead have thresholds $\alpha_1 < \alpha_2$ such that $(U, \alpha) \succeq (\overline{U}, 1-\alpha)$ for $\alpha \leq \alpha_1$, $(\overline{U}, 1-\alpha) \succeq (U, \alpha)$ for $\alpha \geq \alpha_2$, but have no preference for $\alpha \in (\alpha_1, \alpha_2)$.

(R4) Pointwise determinacy of preferences: if $(U_i, \alpha_i) \succeq (V_i, \beta_i)$ for $i = 1, \ldots, n$, then $\{(U_i, \alpha_i) : i = 1, \ldots, n\} \succeq \{(V_i, \beta_i) : i = 1, \ldots, n\}$.

Although this postulate certainly seems reasonable, there are subtleties. For example compare $(U, 1)$ and $(V, 10^{-6})$, where V is an extremely unlikely event, say a lottery win which is a million times the payoff in €. One may reasonably prefer $(V, 10^{-6})$ which gives $(1-10^{-6}) \cdot 1000000 = 999999$€ if $w \in V$ and $(-10^{-6}) \cdot 1000000 = -1$€ otherwise to $(U, 1)$ which is set to always gives 0€. We could argue that even if the chances to win the lottery are quite low, the stake is low, too, so it does not hurt to take the risk. On the other hand, consider the book $\mathcal{B}_n$ consisting of n copies of $(U, 1)$ compared to the book $\mathcal{B}'_n$ consisting of n copies of $(V, 10^{-6})$. According to (R4), $\mathcal{B}'_n \succeq \mathcal{B}_n$. But for n of high dimension, say $n = 10^6$, the agent might not accept to risk loosing $n \cdot 10^{-6} \cdot 1000000$€ for a very small chance to win $n \cdot (1-10^{-6}) \cdot 1000000$€ since the stake is too high, regardless of whether the potential win would also be much higher.

These rationality postulates make it possible to associate with each set $U \subseteq W$ a number α_U, which intuitively is a measure of the likelihood of U (from the perspective of one agent). We set $\alpha_U := \sup\{\alpha : (U, \alpha) \succeq (\overline{U}, 1-\alpha)\}$. This is based on the following insights. It is a consequence of (R1) that $(U, 0) \succeq (\overline{U}, 1)$. As observed earlier, as α grows, (U, α) becomes less attractive and $(\overline{U}, 1-\alpha)$ becomes more attractive. Since again with (R1) we have that $(\overline{U}, 0) \succeq (U, 1)$, we can see that there is some point α^* at which, roughly speaking, (U, α^*) and $(\overline{U}, 1-\alpha^*)$ are in balance (remember that (R3) requires the bets (U, α) and $(\overline{U}, 1-\alpha)$ to be always comparable). We defined α_U to be α^*. Indeed, it is not hard to show that for an agent who is rational in the sense of (R1) – (R4), we have that $(U, \alpha) \succeq (\overline{U}, 1-\alpha)$ for all $\alpha < \alpha_U$ and $(\overline{U}, 1-\alpha) \succeq (U, \alpha)$ for all $\alpha > \alpha_U$. It is not clear, but also not important as the difference would be very slim, what happens at α_U itself; the agent's preferences could be either way. We said that we see α_U as an indicator of how likely the agent deems U. This is because we assume that the more he is willing to bet on U (the larger $\sup\{\alpha : (U, \alpha) \succeq (\overline{U}, 1-\alpha)\}$), the more likely he consideres it to be actually true.

The core insight of Ramsey's argument can now be stated in

Theorem 2.1.1. Consider some $U_1, U_2 \subseteq W$ with $U_1 \cap U_2 = \emptyset$. An agent satisfying our rationality postulates guarantees that $\alpha_{U_1 \cup U_2} = \alpha_{U_1} \cup \alpha_{U_2}$. In particular, the function defined by $\mu(U) = \alpha_U$ is a probability measure.

This is our desired justification for requiring the assignments of numbers to subsets of W to be a probability measure: if an agent behaves rationally in the sense of (R1) – (R4), then he assigns numbers to subsets of W like a probability measure.

Proof. First of all note that $\mu(W) = \alpha_W = 1$ follows directly from the fact that for all α and w we have that $||(W,\alpha)||_w = 1-\alpha \geq \alpha - 1 = ||(\emptyset, 1-\alpha)||_w$, and thus $(W,\alpha) \succeq (\emptyset, 1-\alpha)$, so in particular $\sup\{\alpha : (W,\alpha) \succeq (\emptyset, 1-\alpha)\} = 1$.

To prove that $\alpha_{U_1 \cup U_2} = \alpha_{U_1} + \alpha_{U_2}$ for disjoint U_1 and U_2, we assume the opposite that there are U_1, U_2 with $U_1 \cap U_2 = \emptyset$ but $\alpha_{U_1 \cup U_2} \neq \alpha_{U_1} + \alpha_{U_2}$. We show that this implies that there is a so–called *Dutch book* $\mathcal{B}$ such that $\mathcal{B} \succeq \overline{\mathcal{B}}$. Here, $\mathcal{B}$ is a Dutch book if it has a guaranteed negative payoff: $||\mathcal{B}||_w < 0$ for all $w \in W$, and $\overline{\mathcal{B}}$ is the book comprised of all the complements of bets in $\mathcal{B}$. Since $||(U,\alpha)||_w + ||(\overline{U}, 1-\alpha)||_w = 0$ for all w, which implies $||\mathcal{B}||_w + ||\overline{\mathcal{B}}||_w = 0$, and hence $||\mathcal{B}||_w < ||\overline{\mathcal{B}}||_w$, we could conclude with (R1) that $\overline{\mathcal{B}} \succ \mathcal{B}$, contradicting our assumption.

Towards this, let $\alpha_{U_1} + \alpha_{U_2} < \alpha_{U_1 \cup U_2}$ (the case of ">" follows similarly). There exist $\alpha_1 > \alpha_{U_1}, \alpha_2 > \alpha_{U_2}$, and $\alpha < \alpha_{U_1 \cup U_2}$ with $\alpha_1 + \alpha_2 < \alpha$ (by density of $[0,1]$). Then, by our selection of α's, we have $(U_1, \alpha_1) \prec (\overline{U}_1, 1-\alpha_1)$, $(U_2, \alpha_2) \prec (\overline{U}_2, 1-\alpha_2)$, and $(\overline{U_1 \cup U_2}, 1-\alpha) \prec (U_1 \cup U_2, \alpha)$. Let

$$\mathcal{B} := \{(\overline{U}_1, 1-\alpha_1), (\overline{U}_2, 1-\alpha_2), (U_1 \cup U_2, \alpha)\},$$

so

$$\overline{B} = \{(U_1, \alpha_1), (U_2, \alpha_2), (\overline{U_1 \cup U_2}, 1-\alpha)\}.$$

By (R4), $\mathcal{B} \succ \overline{\mathcal{B}}$. But $\mathcal{B}$ is a Dutch book:

$$||\mathcal{B}||_w = \begin{cases} -1 + \alpha_1 + \alpha_2 + 1 - \alpha, & w \in U_1 \\ \alpha_1 - 1 + \alpha_2 + 1 - \alpha, & w \in U_2 \\ \alpha_1 + \alpha_2 - \alpha, & w \in \overline{U_1 \cup U_2} \end{cases} = \alpha_1 + \alpha_2 - \alpha < 0.$$

□

2.2 Problems with Probabilities for Uncertainty

Desprite its widespread acceptance, using probabilities to represent uncertainty is not without problems. One is illustrated in the following

Example 2.2.1. Suppose there is an urn which contains 100 marbles of which 30 are red and the others are blue or yellow with an unknown distribution. An agent is to place a bet (U_c, α) on a color $c \in \{r, b, y\}$ which gives $1 - \alpha$ if $w = c$ and $-\alpha$ otherwise. Experiments show that people tend to favor red over blue and yellow. However, if they are asked to choose between just blue and yellow, they are indifferent. It is legitimate to interpret this like that they consider blue and yellow equally likely as an outcome. But this would mean they had better chosen blue or yellow instead of red in the first place because of a higher chance of winning $(0.35 > 0.3)$.

We deduce that it is at times rather tricky to consistently assign probabilites to all possible worlds — which is one weakness of this concept. One possible solution is to not consider one probability measure but a set of them, so that the agent is not required to make a definite choice on probabilites of events whose likelihood he does not understand. In the example above, this would mean that the agent's uncertainty about the probabilites can be represented by $\mathcal{P} := \{\mu_a : a \in [0, 0.7], \mu_a(r) = 0.3, \mu_a(b) = a, \mu_a(y) = 0.7 - a\}$. Another way would be to make only some subsets (those that are understood by the agent) measurable. In the example above, $\mathcal{F} := \{\emptyset, \{r\}, \{b, y\}, \{r, b, y\}\}$ would be a good choice.

Now, if we consider a set of probability measures instead of a single one, it becomes unclear to say how probable we consider a certain event. It might be inpractical to always speak about the whole range of probabilities we assign to an event as given by our set of probability measures. A solution to this problem is to concentrate on the lower or upper end of the range. So define, given a set $\mathcal{P}$ of probability measures on W,

$$\text{the } \textit{lower probability} \text{ of } U\text{: } \mathcal{P}_*(U) := \inf\{\mu(U) : \mu \in \mathcal{P}\},$$

and

$$\text{the } \textit{upper probability} \text{ of } U\text{: } \mathcal{P}^*(U) := \sup\{\mu(U) : \mu \in \mathcal{P}\}.$$

In the example we would have $\mathcal{P}_*(r) = \mathcal{P}^*(r) = 0.3, \mathcal{P}_*(b) = \mathcal{P}_*(y) = 0$, and $\mathcal{P}^*(y) = \mathcal{P}^*(y) = 0.7$. Also with the second approach of taking only some subsets to be measurable there is a problem. We might want to give a measurement for a subset that our probability function is not explicitly defined on, despite the uncertainty that goes with this. Again there are two ways of dealing with this issue of extending the probability distribution μ from $\mathcal{F}$ to the entire set $\mathcal{P}(W)$: For arbitrary $U \subseteq W$, define

$$\text{the } \textit{inner measure } \mu_*(U) := \sup\{\mu(V) : V \subseteq U, V \in \mathcal{F}\},$$

and

$$\text{the } \textit{outer measure } \mu^*(U) := \inf\{\mu(V) : V \supseteq U, V \in \mathcal{F}\}.$$

For W finite and $U \in \mathcal{F}$: $\mu^*(U) = \mu_*(U) = \mu(U)$. In general: $\mu_*(U) \leq \mu^*(U)$.

The following theorems establish connections between the lower and upper probabilities and the inner and outer measures.

Theorem 2.2.2. Let μ be a probability measure on a subalgebra $\mathcal{F} \subseteq \mathcal{F}'$ and let $\mathcal{P}_\mu$ be the set of all extensions of μ to $\mathcal{F}'$. Then, $\mu_*(U) = (\mathcal{P}_\mu)_*(U)$ and $\mu^*(U) = (\mathcal{P}_\mu)^*(U)$ for all $U \in \mathcal{F}'$.

Without proof.

Theorem 2.2.3. Consider a preference relation $\succeq$ on books on W, satisfying the rationality postulates (R1) – (R4). Then, there is a set of probability measures $\mathcal{P}$ such that for any $U \subseteq W$ we have that $\mathcal{P}_*(U) = \sup\{\alpha : (U, \alpha) \succeq (\overline{U}, 1 - \alpha)\}$ and $\mathcal{P}^*(U) = \inf\{\alpha : (\overline{U}, 1 - \alpha) \succeq (U, \alpha)\}$.

Without proof.

2.3 Plausibility Measures

In this section we will consider an approach to representing uncertainty which generalizes the approach based on probabilities that we had so far. Instead of probabilities we now use what are called *plausibility measures*. Where a probability measure maps sets in an algebra $\mathcal{F}$ over a set W of worlds to $[0, 1]$, a plausibility measure maps sets in $\mathcal{F}$ to some arbitrary partially ordered set, in which sense it is more general. If Pl is a plausibility measure, Pl(U) denotes the plausibility of U. If $\text{Pl}(U) \leq \text{Pl}(V)$, then V is at least as plausible as U. Because the ordering is partial, it could be that the plausibility of two different sets is incomparable. This means that an agent may not be prepared to order every two sets in terms of plausibility.

Formally, a *plausibility space* is a tuple $S = (W, \mathcal{F}, \text{Pl})$, where W is a set of worlds W, $\mathcal{F}$ is an algebra over W, and Pl maps sets in $\mathcal{F}$ to some set D of *plausibility values* partially ordered by a relation $\leq$ (so that $\leq$ is reflexive, transitive, and antisymmetric). D is assumed to contain two special elements, $\top$ and $\bot$, such that $\bot \leq d \leq \top$ for all $d \in D$. Note that $\leq, \bot$, and $\top$ of course depend on D but we omit a respective index since D is usually clear from the context. Finally, the following three requirements on S must be met:

1. $\text{Pl}(\emptyset) = \bot$.

2. $\text{Pl}(W) = \top$.

3. If $U \subseteq V$, then $\text{Pl}(U) \leq \text{Pl}(V)$.

It is straightforward to verify that by definition, probability measures, lower and upper probabilities, inner and outer measures are instances of plausibility measures, where $D = [0, 1]$, $\bot = 0$, $\top = 1$, and $\leq$ is the standard ordering on the reals. Particularly, in all these examples plausibility values are totally ordered.

To see that the ordering of plausibility values is not always total, consider a set $\mathcal{P}$ of probability measures on W. Both $\mathcal{P}_*$ and $\mathcal{P}^*$ provide a way of comparing the likelihood of two subsets U and V of W. The two ways are incomparable; it is easy to find a set $\mathcal{P}$ of probability measures on W such that $\mathcal{P}_*(U) < \mathcal{P}_*(V)$ but $\mathcal{P}^*(U) > \mathcal{P}^*(V)$. But rather than choosing between $\mathcal{P}_*$ and $\mathcal{P}^*$, it is possible (and sometimes preferable) to associate a different plausibility measure with $\mathcal{P}$ that captures both. Towards this, let $D := \{(a, b) : 0 \leq a \leq b \leq 1\}$ and define $(a, b) \leq (a', b') :\Leftrightarrow a \leq a' \land b \leq b'$. This puts a partial order on D with $\bot = (0, 0)$ and $\top = (1, 1)$. Now, define $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}(U) := (\mathcal{P}_*(U), \mathcal{P}^*(U))$. Clearly, this definition satisfies all the requirements on a plausibility measure, but it puts only a partial order on events.

The problem with only considering lower and upper bounds or even a combination like $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}$ is that they lose information in many situations. Suppose for example that we have an urn with 100 marbles whose color is either red, blue, or yellow. Now, consider the following two experiments:

1. At most 50 marbles are blue and at most 50 marbles are yellow. There is no information about the number of red marbles. Formally, that is $\mathcal{P} = \{\mu : \mu(b), \mu(y) \leq 0.5\}$.

2. There are exactly as many blue as yellow marbles. Again there is no information about the number of red marbles. Formally, that is $\mathcal{P}' = \{\mu : \mu(b) = \mu(y)\}$.

Obviously, $\mathcal{P}' \subsetneq \mathcal{P}$, but $\mathcal{P}_* = \mathcal{P}'_*$ and $\mathcal{P}^* = \mathcal{P}'^*$. So the fact that $\mathcal{P}'$ provides more information than $\mathcal{P}$ is lost when only lower and upper probabilites are considered.

To not lose this kind of information, we define yet another plausibility measure. For this, we let $D_{\mathcal{P}} := \{f : \mathcal{P} \to [0, 1]\}$ be the set of all functions from $\mathcal{P}$ to $[0, 1]$. The standard pointwise ordering on functions, that is, $f \leq g$ if $f(\mu) \leq g(\mu)$ for all $\mu \in \mathcal{P}$, gives a partial order on $D_{\mathcal{P}}$. Clearly, $\top(\mu) = 1$ and $\bot(\mu) = 0$ for all $\mu \in \mathcal{P}$. For $U \subseteq W$, let $f_U : \mathcal{P} \to [0, 1]; \mu \mapsto \mu(U)$, and define the plausibility measure $\mathrm{Pl}_{\mathcal{P}}$ by taking $\mathrm{Pl}_{\mathcal{P}}(U) = f_U$. Then, $\mathrm{Pl}_{\mathcal{P}}(U) \leq \mathrm{Pl}_{\mathcal{P}}(V) \Leftrightarrow \mu(U) \leq \mu(V)$ for all $\mu \in \mathcal{P}$, which is exactly what we wanted. In particular, $\mathrm{Pl}_{\mathcal{P}'} \neq \mathrm{Pl}_{\mathcal{P}}$ since $\mathrm{Pl}_{\mathcal{P}}(U) \leq \mathrm{Pl}_{\mathcal{P}'}(U) \Leftrightarrow f_U \leq f'_U \Leftrightarrow f_U(\mu) \leq f'_U(\mu')$ for all $\mu \in \mathcal{P}, \mu' \in \mathcal{P}'$, and $\mu(U) < \mu'(U)$ for some $\mu \in \mathcal{P}, \mu' \in \mathcal{P}'$.

$\mathrm{Pl}_{\mathcal{P}}$ is indeed a plausibility measure, since $\mathrm{Pl}_{\mathcal{P}}(\emptyset) = f_{\emptyset} = \bot$, $\mathrm{Pl}_{\mathcal{P}}(W) = f_W = \top$, and if $U \subseteq V$, then $\mu(U) \leq \mu(V)$ for all $\mu \in \mathcal{P}$.

To see how this representation works, let us consider again Example 2.2.1, where we had $\mathcal{P} = \{\mu_a : a \in [0, 0.7], \mu_a(r) = 0.3, \mu_a(b) = a, \mu(y) = 0.7 - a\}$. Then, for example

- $\mathrm{Pl}_{\mathcal{P}}(r) = f_r$, where $f_r(\mu_a) = 0.3$,
- $\mathrm{Pl}_{\mathcal{P}}(y) = f_y$, where $f_y(\mu_a) = 0.7 - a$,
- $\mathrm{Pl}_{\mathcal{P}}(\{r, b\}) = f_{\{r,b\}}$, where $f_{\{r,b\}}(\mu_a) = 0.3 + a$.

Chapter 3

A Probabilistic Logic: PCTL

In this chapter, we present an example of logic where probabilities are incorporated directly in its syntax and semantics.

This logic is called *PTCL*, for *probabilistic computational tree logic*, and is based on the compuation tree logics *CTL* and *CTL**. These logics are very important for formal methods in hardware and software verification.

We first recall the defintions of CTL and CTL*. Formulae of these logics are evaluated over transition systems $\mathcal{K} = (V, E, (P_i)_{i \in I})$ with $vE \neq \emptyset$ for all $v \in V$ (we can easily guarantee this via adding a self–loop on nodes supposed to be leaves). A *path* in $\mathcal{K}$ is an infinite sequence $p = v_0 v_1 \ldots \in V^\omega$ with $(w_i, w_{i+1}) \in E$ for all i. Let us first define CTL*. It has *state formulae* φ and *path formulae* π, defined by the mutual induction

$$\begin{aligned} \varphi &::= P_i \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{E}\pi \mid \mathbf{A}\pi, \\ \pi &::= \varphi \mid \neg\pi \mid \pi \wedge \pi \mid \mathcal{X}\pi \mid \pi\mathcal{U}\pi. \end{aligned}$$

For all paths $p = v_0 v_1 \ldots$, let $p[i] = v_i v_{i+1} \ldots$ Then, we define the semantics like

- $\mathcal{K}, v \models \mathbf{E}\pi :\Leftrightarrow$ there is a path p starting at v with $\mathcal{K}, p \models \pi$,

- $\mathcal{K}, v \models \mathbf{A}\pi :\Leftrightarrow$ all paths p starting at v satisfy $\mathcal{K}, p \models \pi$,

- $\mathcal{K}, p \models \varphi :\Leftrightarrow \mathcal{K}, v_0 \models \varphi$, where $p = v_0 v_1 \ldots$,

- $\mathcal{K}, p \models \mathcal{X}\pi :\Leftrightarrow \mathcal{K}, p[1] \models \pi$,

- $\mathcal{K}, p \models \pi_1 \mathcal{U} \pi_2 :\Leftrightarrow$ there exists an $i \geq 0$ such that $\mathcal{K}, p[i] \models \pi_2$ and for all $j \leq i$ we have that $\mathcal{K}, p[j] \models \pi_1$.

Now, CTL is the restriction of CTL* admitting only path formulae of the form $\mathcal{X}\varphi$ or $\varphi\mathcal{U}\psi$, where φ, ψ are state formulae. Equivalently we can think of CTL as defined by

$$\varphi ::= P_i \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbf{E}\mathcal{X}\varphi \mid \mathbf{A}\mathcal{X}\varphi \mid \mathbf{E}(\varphi\mathcal{U}\varphi) \mid \mathbf{A}(\varphi\mathcal{U}\varphi).$$

We write $\mathcal{F}\varphi$ for $1\mathcal{U}\varphi$ and $\mathcal{G}\varphi$ for $\neg\mathcal{F}\neg\varphi$. A few exemplary formulae are

- $\mathbf{E}\mathcal{F}\varphi$ — a state where φ holds is reachable (from the current state),
- $\mathbf{A}\mathcal{G}\neg(\varphi \wedge \psi)$ — in all reachable states, φ and ψ exlude each other,
- $\mathbf{A}\mathcal{G}\mathbf{A}\mathcal{F}\varphi$ — on all paths, φ holds infinitely often.

Further, we want to name a couple of properties:

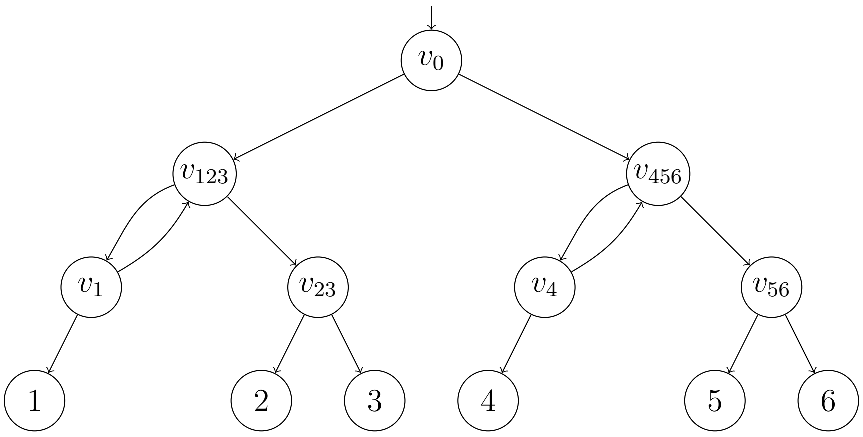
- CTL and CTL* are invariant under bisimulation.
- CTL admits efficient model checking. That is, given $\mathcal{K}, \varphi$, it can be decided in time $\mathcal{O}(||\mathcal{K}|| \cdot |\varphi|)$ at which states of $\mathcal{K}$ the formula φ is true. On the other hand, model checking for CTL* is PSPACE–complete.
- CTL and CTL* both have the finite model property, that is, if a CTL– or a CTL*–formula has any model, then it also has a finite one.
- Satisfiability for CTL is in EXPTIME.

Now, we can introduce the probabilistic variant PCTL with the following syntax:

$$\varphi ::= P_i \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbb{P}_J(\pi), \text{ where } J \subseteq [0,1] \text{ has rational bounds,}$$
$$\pi ::= \mathcal{X}\varphi \mid \varphi\mathcal{U}\varphi \mid \varphi\mathcal{U}^{\leq n}\varphi \text{ for } n \in \mathbb{N}.$$

Sometimes we write things like $\mathbb{P}_{>0.5}$ for $\mathbb{P}_{(0.5,1]}$ or $\mathbb{P}_{=1}$ for $\mathbb{P}_{[1,1]}$, etc. Semantically, we define $\mathcal{K}, p \models \varphi\mathcal{U}^{\leq n}\psi$ if there is an $i \leq n$ such that $\mathcal{K}, p[i] \models \psi$ and $\mathcal{K}, p[j] \models \varphi$ for all $j = 0, \ldots, i-1$. To define the meaning of $\mathbb{P}_J(\pi)$, we introduce the concept of *Markov chains*. These are structures $\mathcal{K} = (V, \Delta, i, (P_i)_{i\in I})$, where $\Delta\colon V \times V \to [0,1]$ is a transition path function such that for all v we have that $\sum_{w\in V} \Delta(v,w) = 1$, and $i\colon V \to [0,1]$ with $\sum_{v\in V} i(v) = 1$ is the initial distribution.

Example 3.0.1. Simulation of a dice by a fair coin. Each edge has probability 1/2. A leaf $n = 1, \ldots, 6$ is reached with probability exactly 1/6, so we have $\mathcal{K}, v_0 \models \mathbb{P}_{1/6}[\mathcal{F}n]$.



Let $Paths(\mathcal{K})$ be the set of all infinite paths $v_0 v_1 \dots$ such that $\Delta(v_i, v_{i+1}) > 0$ for all i. Furthermore, $FinPaths(\mathcal{K})$ is the set of all finite paths, and for an $f \in FinPaths(\mathcal{K})$ let

$$\sigma(f) := \{p \in Paths(\mathcal{K}) \mid f \leq p\}$$

be the set of infinite paths that are identical to f up to its length. We associate with $\mathcal{K}$ the σ–algebra which is the smallest to contain $\sigma(f)$ for every $f \in FinPaths(\mathcal{K})$. Now, define a unique probability measure on this σ–algebra:

$$\mu(\sigma(v_0 \dots v_n)) := i(v_0) \cdot \prod_{0 \leq j < n} \Delta(v_j, v_{j+1}).$$

Lemma 3.0.2 (Measurability Lemma)**.** For every PCTL path formula π and every state v of a Markov chain $\mathcal{K}$, the set

$$Paths(v, \pi) := \{p \in Paths(v) \mid \mathcal{K}, p \models \pi\}$$

is measurable, where $Paths(v) := \{p \in Paths(\mathcal{K}) : p \text{ starts at } v\}$.

Proof. $Paths(v, \pi)$ is a countable union of sets $\sigma(f)$ for $f \in FinPaths(\mathcal{K})$. To see this, we consider the following three cases.

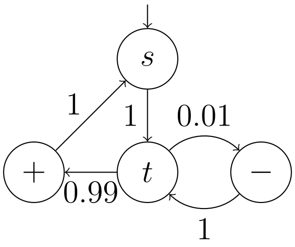
- $\pi = \mathcal{X}\varphi$: $Paths(v, \pi) = \bigcup\{\sigma(vw) : \mathcal{K}, w \models \varphi\}$.
- $\pi = \varphi\mathcal{U}^{\leq n}\psi$: $Paths(v, \pi) = \bigcup\{\sigma(v_0 \dots v_i) : i \leq n, \mathcal{K}, v_i \models \psi, \mathcal{K}, v_j \models \varphi$ for $j = 0, \dots, i-1$, $v_0 = v\}$.
- $\pi = \varphi\mathcal{U}\psi$: $Paths(v, \pi) = \bigcup_{n \geq 0} Paths(v, \varphi\mathcal{U}^{\leq n}\psi)$.

□

With this, we can define the semantics for $\mathbb{P}_J(\pi)$ as follows:

$$\mathcal{K}, v \models \mathbb{P}_J(\pi) :\Leftrightarrow \mu(Paths(v, \pi)) \in J.$$

Example 3.0.3. A Markov chain $\mathcal{K}$ with $\mathcal{K}, s \not\models \mathbf{A}\mathcal{F}(+)$ but $\mathcal{K}, s \models \mathbb{P}_{=1}\mathcal{F}(+)$.



Without giving a proof, we want to note that model checking for PCTL, that is, given a finite Markov chain $\mathcal{K}$ and a state formula $\varphi \in$ PCTL deciding whether $\mathcal{K}, v \models \varphi$, can be decided in time $||\mathcal{K}||^{\mathcal{O}(1)} \cdot n_{max} \cdot |\varphi|$. Here, n_{max} is the maximal n such that φ contains a subformula $\varphi_1 \mathcal{U}^{\leq n} \varphi_2$.

3.1 QPCTL

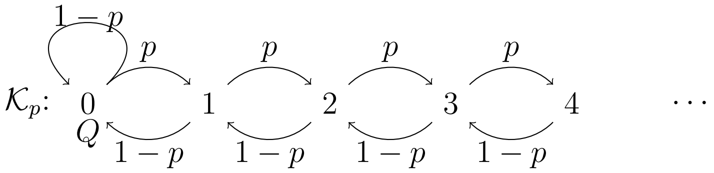
We want to introduce the so–called *qualitative fragment QPCTL* of PCTL, defined as

$$\begin{aligned} \varphi &::= P_i \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathbb{P}_{>0}(\pi) \mid \mathbb{P}_{=1}(\pi), \\ \pi &::= \mathcal{X}\varphi \mid \varphi\mathcal{U}\varphi. \end{aligned}$$

Let us investigate the relationship of QPCTL to CTL. Towards this, we associate with each Markov chain $\mathcal{K}$ the transition system $Q\mathcal{K} := (V, E, (P_i)_{i \in I})$ with $E := \{(v, w) : \Delta(v, w) > 0\}$. For a $\varphi \in$ PCTL and a $\varphi' \in$ CTL, we say $\varphi \equiv \varphi'$ if, and only if, for all Markov chains and their states $\mathcal{K}, v$, we have that $\mathcal{K}, v \models \varphi \Leftrightarrow Q\mathcal{K}, v \models \varphi'$. To give an example, remember that as illustrated in Example 3.0.3, $Q\mathcal{K}, v \models \mathbf{A}\mathcal{F}\varphi \Rightarrow \mathcal{K}, v \models \mathbb{P}_{=1}(\mathcal{F}\varphi)$, but the converse does not hold, so $\mathbf{A}\mathcal{F}\varphi \not\equiv \mathbb{P}_{=1}(\mathcal{F}\varphi)$. With the following lemma, we further generalize this and show that QPCTL is not contained in CTL.

Lemma 3.1.1. There exists no CTL–formula that is equivalent to $\mathbb{P}_{=1}(\mathcal{F}\varphi)$ or to $\mathbb{P}_{>0}(\mathcal{G}\varphi)$, even for atomic φ.

Proof. To see this, assume that $\varphi' \in$ CTL is equivalent to $\mathbb{P}_{=0}(\mathcal{F}Q)$, where Q is an atomic formula. Now, consider for $p \in [0, 1]$ the following infinite transition system:



It is not hard to see that for $p \leq 1/2$, state 0 is almost surely reached infinitely often, whereas for $p > 1/2$ the chain moves to the right:

$$\mathcal{K}, n \models \mathbb{P}_{<1}(\mathcal{F}Q) \text{ and } \mathcal{K}, n \models \mathbb{P}_{=0}(\mathcal{G}\mathcal{F}Q) \text{ for all } n > 0.$$

Thus,

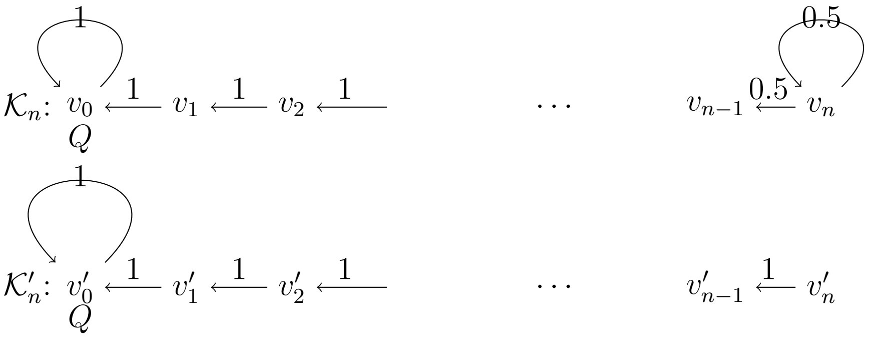
$$\mathcal{K}_{1/4}, 1 \models \mathbb{P}_{=1}(\mathcal{F}Q) \text{ and } \mathcal{K}_{3/4}, 1 \models \neg\mathbb{P}_{n=1}(\mathcal{F}Q), \text{but } Q\mathcal{K}_{1/4} = Q\mathcal{K}_{3/4}.$$

□

On the other hand, CTL is also not contained in QPCTL:

Lemma 3.1.2. No QPCTL–formula is equivalent to $\mathbf{A}\mathcal{F}\varphi$ or to $\mathbf{E}\mathcal{G}\varphi$, even for atomic φ.

Proof. Consider the following transition systems:



We have that

$$\mathcal{K}'_n, v'_n \models \mathbf{A}\mathcal{F}Q \text{ and } \mathcal{K}_n, v_n \not\models \mathbf{A}\mathcal{F}Q.$$

Define for $\varphi \in$ QPCTL the nesting depth $d(\varphi)$ of $\mathbb{P}_J$–operators in φ. If $d(\varphi) < n$, then $\mathcal{K}_n, v_n \models \varphi \Leftrightarrow \mathcal{K}'_n, v'_n \models \varphi$. Hence, $\varphi \not\equiv \mathbf{A}\mathcal{F}Q$. □

Chapter 4

Logics for Dependence, Independence, and Imperfect Information

All the concepts with respect to uncertainty we introduced so far were about representing the uncertainty itself (e.g. by assigning probabilites to events according to how likely we consider them to occur). In contrast, this chapter logically formalizes not uncertainty itself but rather a feature of situations that is often responsible for whether there is uncertainty in the first place — and what kind of uncertainty — or not.

What we are referring to is the notion of dependence and independence. If two events A and B (completely) depend on each other, then from certainty about A we gain certainty also about B. On the other hand, if A and B are (completely) independent, then the uncertainty about B will remain on the exact same level even in the wake of gaining complete information about A. When we want to capture logically the ideas of dependence and independence we quickly realize that this is possible not without some new ideas. Let us illustrate the problem behind this with an example: Compare the statements "y depends on x" or "x and y are independent" to "x divides y". To reason about the latter we fix a structure $\mathfrak{A}$ where divisibility is well–defined, and an assignment $s\colon \{x, y\} \to A$. Now, we determine whether $\mathfrak{A} \models_s$ "x divides y". On the other hand, dependence and independence do not manifest themselves for a single assignment but only in larger amounts of data. So we need alternatives to the normal Tarski–style $\mathfrak{A} \models_s$"x and y are independent" for a reasonable evaluation of such formulae speaking about dependence or independence. This chapter is dedicated to analyzing a couple of those alternatives.

4.1 Henkin Quantifiers and Independence Friendly Logic

The idea of *Henkin quantifiers* (also called *branching quantifiers*) is the following. Consider the FO–formula

$$\forall x \exists v \forall y \exists u \varphi(x, y, u, v).$$

Clearly, the choice of v might depend on u. The Henkin quantifier prevents this and makes v and u independent:

$$\begin{pmatrix} \forall x \exists u \\ \forall y \exists v \end{pmatrix} \varphi(x, y, u, v).$$

Its semantics can be defined quite simply with Skolem functions:

$$\mathcal{A} \models \begin{pmatrix} \forall x \exists u \\ \forall y \exists v \end{pmatrix} \varphi(x, y, u, v) :\Leftrightarrow \text{there exist } f, g \colon A \to A \text{ such that}$$
$$(\mathcal{A}, f, g) \models \forall x \forall y \varphi(x, fx, y, gy).$$

Alternatively, we can base the semantics of the Henkin quantifier on a model checking game with imperfect information (as opposed to model checking games for normal FO, where both players have perfect information). In the model checking game for $\mathfrak{A}$ and $(\forall x \exists u \forall y \exists v)\ \varphi(x, y, u, v)$, Falsifier choses $x \mapsto a$, Verifier choses $u \mapsto b$, Falsifier choses $y \mapsto c$, and Verifier choses $v \mapsto d$. Here, the choice of d may depend on c only, not on a or b. This is why we say the game has imperfect information. We imagine Verifier having no information about the choices of Falsifier regarding x and y.

What might be counterintuitive: this requirement actually increases the expressive power of the logic. In classical FO it is always efficiently decidable (in LOGSPACE) whether a given finite structure is a model of φ. On the other hand, consider the Henkin quantifier

$$\begin{pmatrix} \forall x \exists u \\ \forall y \exists v \end{pmatrix} (x = y \to v = u) \wedge (u = 0 \vee u = 1 \vee u = 2) \wedge (Exy \to u \neq v).$$

It is not hard to see that for a graph $G = (V, E)$ and three distinct nodes $0, 1, 2 \in V$, we have that $G, 0, 1, 2$ is a model of this quantifier if, and only if, G is 3-colourable. Thus, it defines an NP-complete problem.

Another approach to capture dependence and independence by extending FO with a new quantifier is the so–called *independence friendly logic* (IF-logic). If φ is a IF-formula, x a variable, and W a finite set of variables, then also $(\exists x/W)\varphi$ and $(\forall x/W)\varphi$ are IF-formulae. In these quantifiers, the variable x is supposed to be independent from those in W.

Let us first define its semantics in terms of game-theory. In the evaluation game for $(Qx/W)\varphi$, the value for x must be chosen independently from the values of variables in W. We realize this formally by a contraint on strategies requiring that at two positions $(Qx\varphi, s)$ and $(Qx\varphi, s')$ with $s(y) = s'(y)$ for all $y \in W$ the same values for x must be chosen. The rest is like the normal model checking game: $\varphi(\overline{a})$ is true if Verifier has a winning strategy.

Also with this logic it is possible to define NP–complete problems. For example consider

$$\forall x \forall y (\exists u/\{y\})(\exists v/\{x,u\})((x = y \to u = v) \land (u = y \to v = x) \land Exu).$$

A graph $G = (V, E)$ admits a perfect matching if, and only if, it is satisfies this formula.

We can define the semantics of (Qx/W) also by replacing existential quantifiers $(\exists x/W)$ with Skolem functions where arguments are the variables outside W. The above formula for perfect matching then becomes

$$\exists f \exists g \forall x \forall y (x = y \to fx = gy) \land (fx = y \to gy = x) \land Exfx,$$

which is equivalent to

$$\exists f \forall x (ffx = x \land Exfx).$$

Without proof we state

Theorem 4.1.1. IF–logic $\equiv \Sigma^1_1$.

4.2 Dependence and Independence as Atomic Statements

There is a different approach which, rather than stating dependencies as annotations of quantifiers, treats them as atomic statements. To cope with the fact that dependence and independence are concepts that speak about multiple and not singular data, the model–theoretic semantics is then in form of sets of assignments instead of a single one. Sets of assignments are called *teams*.

Assignments are usually denoted by $s\colon V \to A$, where V is a set of variables and A is the universe of a structure $\mathfrak{A}$. Teams are denoted by X. All assignments in a team have the same domain and codomain. Note that

V can be empty. There are two teams with domain $\emptyset$, namely $X = \emptyset$ and $X = \{\emptyset\}$.

A *dependence atom* is a term $=(\overline{x}, y)$ which means "y is determined by $\overline{x}$", or "y is functionally dependend on $\overline{x}$". For a team X of assignments $s\colon V \to A$, where $V = \{x_1, \ldots, x_n, y\}$, this means

$$\mathfrak{A} \models_X =(\overline{x}, y) :\Leftrightarrow \bigwedge_{i=1}^{n} s(x_i) = s'(x_i) \Rightarrow s(y) = s'(y) \text{ for all } s, s' \in X.$$

Important is also the expanded form of dependence atoms $=(\overline{x}, \overline{y})$, saying that all variables from $\overline{y}$ depend on $\overline{x}$. With $=(y)$ we express that the value of y is constant in X.

An *independence atom* $x \bot y$ for us is not just the negation of a dependence atom. Rather than stating that two events are not dependend, it is supposed to express that two events are *completely* independent in the sense that every possible pattern of values for (x, y) occurs. Formally, this means

$$\mathfrak{A} \models_X x \bot y \Leftrightarrow (\forall s, s' \in X)(\exists s'' \in X)(s''(x) = s(x) \wedge s''(y) = s'(y)).$$

Let us illustrate this definition with an example.

Example 4.2.1. Suppose you know X and that some assignment s is in X. Now, you want to gather information about $s(y)$. If you are told $s(x)$ with $x \bot y$, then you cannot infer anything about $s(y)$. Indeed, for all potential values $a \in \{a : (\exists s' \in X) s'(y) = a\}$ of $s(y)$, there is by definition of $x \bot y$ a $s'' \in X$ with $s''(x) = s(x)$ and $s''(y) = a$.

A special case of independence is when one of the variables is a constant as in the following two examples. In fact, it is not hard to see that a constant is independent from every other variable including itself: $=(x) \equiv x \bot x$.

Example 4.2.2. Galileo: The time of descent is independent from the mass.

Example 4.2.3. Einstein: The speed of light is independent from the observer's state of motion.

Like with dependence atoms, we can naturally generalize independence atoms to speak about tuples of variables, written as $\overline{x} \bot \overline{y}$. Let us generalize this even further to

$$\begin{aligned} \overline{x} \bot_{\overline{z}} \overline{y} :\Leftrightarrow{} & (\forall s, s' \in X) s(\overline{z}) = s'(\overline{z}) \\ & \to (\exists s'' \in X) s''(\overline{z}) = s(\overline{z}) = s'(\overline{z}) \wedge s''(\overline{x}) = s(\overline{x}) \wedge s''(\overline{y}) = s'(\overline{y}). \end{aligned}$$

Directly from their definition, we have the following relationships between dependence atoms and independence atoms regarding their expressiveness:

- $=(\overline{z}, \overline{x}) \Rightarrow \overline{x} \bot_{\overline{z}} \overline{y}$ (select $s'' := s'$).
- $\overline{x} \bot_{\overline{z}} \overline{y} \Rightarrow =(\overline{z}, \overline{x} \cap \overline{y})$ $(s(\overline{x} \cap \overline{y}) = s''(\overline{x} \cap \overline{y}) = s'(\overline{x} \cap \overline{y}))$.
- It follows that $=(\overline{z}, \overline{x}) \Leftrightarrow \overline{x} \bot_{\overline{z}} \overline{x}$.

So dependence is a special case of the general form of independence.

Besides dependence and independence, there are other atomic properties of teams worth investigating:

- Inclusion $\overline{x} \subseteq \overline{y}$ with
$$\mathfrak{A} \models_X \overline{x} \subseteq \overline{y} :\Leftrightarrow (\forall s \in X)(\exists s' \in X) s(\overline{x}) = s'(\overline{y}).$$
- Exclusion $\overline{x}|\overline{y}$ with
$$\mathfrak{A} \models_X \overline{x}|\overline{y} :\Leftrightarrow (\forall s, s' \in X) s(\overline{x}) \neq s'(\overline{y}).$$

Also, $=(\overline{x}, \overline{y})$ and $\overline{x}|\overline{y}$ are downward–closed:

$$\mathfrak{A} \models_X \varphi, Y \subseteq X \Rightarrow \mathfrak{A} \models_Y \varphi,$$

whereas $\overline{x} \subseteq \overline{y}$ are not downwards closed, but closed under unions of teams:

$$\text{If } X = \bigcup_{i \in I} X_i \text{ and } \mathfrak{A} \models_{X_i} \overline{x} \subseteq \overline{y} \text{ for all } i \in I, \text{ then } \mathfrak{A} \models_X \overline{x} \subseteq \overline{y}.$$

4.3 Team Semantics for Logics of Dependence and Independence

Now, we add logical operators $\vee, \wedge, \exists, \forall$ to the various forms of dependency atoms to obtain full-fledged logics for reasoning about dependence and independence:

- Dependence logic: FO + $=(\overline{x}, \overline{y})$.
- Independence logic: FO + $\overline{x} \bot_{\overline{z}} \overline{y}$.
- Inclusion logic: FO + $\overline{x} \subseteq \overline{y}$.

We always assume formulae of these logics to be in NNF such that negations are applied only to FO–atoms. It is still necessary to extend the definition of team semantics to FO.

Our goal is to have

$$\mathfrak{A} \models_X \varphi \Leftrightarrow \mathfrak{A} \models_s \varphi \text{ for all } s \in X.$$

Thus, we define the semantics of a formula φ inductively as follows.

- $\mathfrak{A} \models_X \varphi :\Leftrightarrow \mathfrak{A} \models_s \varphi$ for all $s \in X$, if φ is an FO–literal.
- $\mathfrak{A} \models_X \varphi = \psi \wedge \vartheta :\Leftrightarrow \mathfrak{A} \models_X \psi$ and $\mathfrak{A} \models_X \vartheta$.
- $\mathfrak{A} \models_X \varphi = \psi \vee \vartheta :\Leftrightarrow \exists Y \exists Z$ with $X = Y \cup Z$ and $\mathfrak{A} \models_Y \psi$, $\mathfrak{A} \models_Z \vartheta$.

  Observe that this definition causes unusual behaviour, for example we have that $=\!(x) \not\equiv =\!(x) \vee =\!(x)$. But is is certainly the one consistent with our semantic goal.

  It also allows us to express NP–complete problems such as 3-SAT. Towards this, with a propositional formula $\varphi = \bigwedge_{i=1}^{n}(X_{i_1} \vee X_{i_2} \vee X_{i_3})$ we associate the team

  $$\begin{aligned} Z_\varphi := \{&(\mathit{clause}, \mathit{position}, \mathit{variable}, \mathit{parity}) \mapsto (i, j, X, \sigma) : \\ &\text{in clause } i \text{ at position } j, \text{ the variable } X \text{ occurs with parity } \sigma\} \end{aligned}$$

  with domain $\{(\mathit{clause}, \mathit{position}, \mathit{variable}, \mathit{parity})\}$ and codomain as induced by φ. We claim that φ is satisfiable if, and only if,

  $$\mathfrak{A} \models_{Z_\varphi} =\!(\mathit{clause}, \mathit{position}) \vee =\!(\mathit{clause}, \mathit{position}) \vee =\!(\mathit{variable}, \mathit{parity}),$$

  where $\mathfrak{A}$ is just $(\mathit{codomain}(Z_\varphi))$. Now, the latter holds if, and only if, $Z_\varphi = Z_1 \cup Z_2 \cup Z_3$ such that

  1. $\mathfrak{A} \models_{Z_1} =\!(\mathit{clause}, \mathit{position})$,
  2. $\mathfrak{A} \models_{Z_2} =\!(\mathit{clause}, \mathit{position})$,
  3. $\mathfrak{A} \models_{Z_3} =\!(\mathit{variable}, \mathit{parity})$.

  With this, the only requirement on Z_1 and Z_2 is that both contain at most one variable per clause, so that Z_3 must contain at least one variable for each clause. Each time a particular variable occurs in Z_3, it has to occur with the same parity, so we can use it to make the respective clauses true.

- $\mathfrak{A} \models_X \varphi = \forall y \psi :\Leftrightarrow \mathfrak{A} \models_{X[y \mapsto A]} \psi$. Here, $X[y \mapsto A]$ is defined as the set $\{s[y \mapsto a] : s \in X, a \in A\}$, where $s[y \mapsto a]$ denotes the extension/update of s by mapping y to a.

- $\mathfrak{A} \models_X \varphi = \exists y \psi :\Leftrightarrow$ there is a function $F\colon X \to \mathcal{P}(A) \setminus \{\emptyset\}$ such that $\mathfrak{A} \models_{X[y \mapsto F]} \psi$, where $X[y \mapsto F] := \{s[y \mapsto a] : s \in X, a \in F(s)\}$.

  One might legitimately ask why not say that $\mathfrak{A} \models_X \exists y \psi :\Leftrightarrow$ there is a function $F\colon X \to A$ such that $\mathfrak{A} \models_{\{s[y \mapsto F(s)] : s \in X\}} \psi$. In fact, there is no difference between these two ways (we call the first one *lax* semantics

> and the second one *strict* semantics) for FO and also for dependence logic. But for stronger logics, only lax semantics works the way we want it to.
>
> For example consider $\exists z(x \subseteq z \wedge y \subseteq z)$. In lax semantics, this formula is a tautology, since we are allowed to extend x and y by different values so that we can set $F(s) := \{s(x), s(y)\}$. On the other hand, under strict semantics φ says something else. Consider the team X consisting of the two assignments $(x \mapsto 1, y \mapsto 2, u \mapsto 1)$ and $(x \mapsto 1, y \mapsto 2, u \mapsto 2)$. It is easy to see that this team induces a model of our formula since also here we can make two different selections. Now, consider $X \upharpoonright \{x, y\}$ which is $\{(x \mapsto 1, y \mapsto 2)\}$. Clearly, this is not a model any more. Thus, strict semantics violates the locality principle that the meaning of a formula only depends on the variables free in it, which is why we deem it not the right choice.

With lax semantics, we can express problems that are even beyond $\mathcal{L}_{\infty,\omega}$. For example $(A, <) \models \exists x \exists y(y < x \wedge y \subseteq x)$ is true if, and only if, $(A, <)$ contains an infinite descending chain. To see this, note that the formula is satisfied exactly if there is a team X with domain $\{x, y\}$ such that $(A, <) \models_X y < x \wedge y \subseteq x$. So X is a set of assignments $s\colon (x, y) \mapsto (a, b)$ such that $b < a$ and there is an assignment $s' \in X$ with $s'\colon (x, y) \mapsto (b, c)$. Now, the same must hold for s', and so on ad infinitum.

Note that all these logics have the empty set property: $\mathfrak{A} \models_\emptyset \varphi$ for all φ. Further it is easy to see that the properties of being downwards closed, or closed under union of teas is preserved by the semantic rules of all logical operators. Hence the fact that dependence and exclusion atoms are downwards closed extends to all formulae of dependence and exclusion logic. Similarly the fact that inclusion atoms are closed under unions of teams extends to all formulae of inclusion logic.

For pure FO–formulae we have the following *flatness*–property:

$$\mathfrak{A} \models_X \varphi \Leftrightarrow \mathfrak{A} \models_{\{s\}} \varphi \text{ for all } s \in X \Leftrightarrow \mathfrak{A} \models_s \varphi \text{ for all } s \in X.$$

4.4 From Team Semantics to Tarski Semantics

To understand the expressive power of the introduced logics with team semantics, we want to express them with classical Tarski semantics of which we already have extensive knowledge with regards to expressiveness.

For sentences φ we simply put $\mathfrak{A} \models \varphi :\Leftrightarrow \mathfrak{A} \models_{\{\emptyset\}} \varphi$. For formulae $\varphi(\overline{x})$ with free variables we have to represent a team with classical methods. What we will do is to identify with a team X of assignments $s\colon \{x_1, \dots, x_k\} \to A$ a k–ary relation $rel(X) := \{(s(x_1), \dots, s(x_k)) : s \in X\} \subseteq A^k$. Then, if $\varphi(\overline{x})$ has vocabulary τ, we will translate it to a formula φ^* of vocabulary $\tau \uplus \{X\}$ that can be evaluated with Tarski semantics, as captured in

Theorem 4.4.1. For every formula $\varphi(\overline{x})$ of dependence/independence/exclusion /inclusion logic there is a sentence $\varphi^*(X)$ of vocabulary $\tau \uplus \{X\}$ in Σ^1_1 such that $\mathfrak{A} \models_X \varphi(\overline{x}) \Leftrightarrow (\mathfrak{A}, X) \models \varphi^*(x)$.

Proof. Let $\varphi(\overline{x})$ be one of the different kinds ot dependence atoms. Further, define $(i) := i_1, \dots, i_k$ and $\overline{x}_{(i)} := (x_{i_1}, \dots, x_{i_k})$.

- $(\overline{x}_{(i)}, \overline{x}_{(j)}) \rightsquigarrow \forall\overline{x}\forall\overline{y}(X\overline{x} \wedge X\overline{y} \wedge \overline{x}_{(i)} = \overline{y}_{(i)} \to \overline{x}_{(j)} = \overline{y}_{(j)})$.
- $\overline{x}_{(i)}|\overline{x}_{(j)} \rightsquigarrow \forall\overline{x}\forall\overline{y}(X\overline{x} \wedge X\overline{y} \to \overline{x}_{(i)} \neq \overline{y}_{(j)})$.
- $\overline{x}_{(i)} \subseteq \overline{y}_{(j)} \rightsquigarrow \forall\overline{x}(X\overline{x} \to \exists\overline{y}(X\overline{y} \wedge \overline{x}_{(i)} = \overline{y}_{(j)}))$.
- $\overline{x}_{(i)} \bot_{\overline{x}(k)} \overline{x}_{(j)} \rightsquigarrow \forall\overline{x}\forall\overline{y}(X\overline{x} \wedge X\overline{y} \wedge \overline{x}_{(k)} = \overline{y}_{(k)}$
  $\to \exists\overline{z}(X\overline{z} \wedge \overline{x}_{(k)} = \overline{z}_{(k)} \wedge \overline{x}_{(i)} = \overline{z}_{(i)} \wedge \overline{y}_{(j)} = \overline{z}_{(j)}))$.
- $\alpha(\overline{x})$ FO–literal $\rightsquigarrow \forall\overline{x}(X\overline{x} \to \alpha(\overline{x}))$.
- $\psi(\overline{x}) \wedge \vartheta(\overline{x}) \rightsquigarrow \psi^*(X) \wedge \vartheta^*(X)$.
- $\psi(\overline{x}) \vee \vartheta(\overline{x}) \rightsquigarrow \exists Y \exists Z(\forall\overline{x}(X\overline{x} \leftrightarrow Y\overline{x} \vee Z\overline{x}) \wedge \psi^*(Y) \wedge \vartheta^*(Z))$.
- $\forall y\psi(\overline{x}, y) \rightsquigarrow \exists Y(\forall\overline{x}\forall y(X\overline{x} \leftrightarrow Y\overline{x}y) \wedge \psi^*(Y))$.
- $\exists y\psi(\overline{x}, y) \rightsquigarrow \exists Y(\forall\overline{x}(X\overline{x} \leftrightarrow \exists yY\overline{x}y) \wedge \psi^*(Y))$.

□

This theorem shows that all those logics are contained in Σ^1_1. We want to be more precise about dependence and inclusion logic. Remember that those are closed under subteams. On the other hand, the corresponding statement for Σ^1_1 is clearly not true, i. e., for $\varphi^* \in \Sigma^1_1$ in general we not have that $(\mathfrak{A}, X) \models \varphi^*$ implies $(\mathfrak{A}, Y) \models \varphi^*$ for every $Y \subseteq X$. Thus, dependence and inclusion logic are strict fragments of Σ^1_1. We will see that we can easily give a more concise characterization of this fragment.

Theorem 4.4.2. For every $\varphi(\overline{x})$ of dependence and inclusion logic there is a sentence $\varphi^*(X) \in \Sigma^1_1$ in which the predicate X for the team occurs only negatively such that $\mathfrak{A} \models_X \varphi(\overline{x}) \Leftrightarrow (\mathfrak{A}, X) \models \varphi^*$.

Proof. All we have to do is to convince ourselves that we can replace "$\leftrightarrow$" by "$\rightarrow$" in all formulae $\varphi^*(X)$ from the previous proof except those for independence and exlusion atoms. But this follows immediately from the fact that dependence and inclusion logic are downward–closed. □

4.5 Model Checking Games for Logics with Team Semantics

Recall the classical model checking games for FO, where for a formula φ in NNF we play a reachability game $\mathcal{G} = (\mathfrak{A}, \varphi) = (V, V_0, V_1, T, E, I)$, where V_0 are positions of Player 0 (Verifier), V_1 are positions of Player 1 (Falsifier), T are terminal positions (those with no outgoing edges), $V = V_0 \uplus V_1 \uplus T$, $E \subseteq V \setminus T \times V$, and $V = \{(\psi, s) : \psi \text{ subformula of } \varphi,\ s\colon \mathit{free}(\psi) \to A\}$. I is the set of initial positions; here: $I = \{(\varphi, \emptyset)\}$.

We have $\mathfrak{A} \models \varphi \Leftrightarrow$ Player 0 has a winning strategy for $\mathcal{G}(\mathfrak{A}, \varphi)$. At a position $v \in T$, Player 0 has won if $v \in \mathit{Win} \subseteq T$, Player 1 has won if $v \in T \setminus \mathit{Win}$. Here, $\mathit{Win} = \{(\psi, s) \in T \mid W \models_s \psi\}$.

Now, different from the common approach, we think of a strategy for Player 0 as a subgraph $S = (W, F) \subseteq (V, E)$ such that

1. $v \in W \cap V_0 \Rightarrow vF \neq \emptyset$,
2. $v \in W \cap V_1 \Rightarrow vF = vE$,
3. $F \subseteq W \times W \cap E$.

$S = (W, F)$ wins from $X \subseteq I$ if $X \subseteq W$ and all players that start at some $v \in X$ and are consistent with S reach a winning position $w \in \mathit{Win}$.

We know that given a reachability game $(\mathcal{G}, \mathit{Win})$ on a finite game graph, one can compute in linear time $\mathcal{O}(|V| + |E|)$

- the maximal subset $X \subseteq I$ from which Player 0 can win, and
- a winning strategy from X.

Let us first introduce model checking games for logics with team semantics that use a reachability condition. That is, we want to reduce the model checking problem $\mathfrak{A} \models_X \varphi$ to a reachability game $\mathcal{G}(\mathfrak{A}, X, \varphi)$. Positions are (ψ, Y), where Y is a team of assignments $s\colon \mathit{free}(\psi) \to A$. Player 1 moves from $(\psi_1 \wedge \psi_2, Y)$ to $(\psi_i, Y \upharpoonright \mathit{free}(\psi_i))$ for an $i \in \{1, 2\}$, and from $(\forall x\psi, Y)$ deterministically to $(\psi, Y[x \mapsto A])$. On the other hand, Player 0 moves from $(\exists x\psi, Y)$ to a position $(\psi, Y[x \mapsto F])$ for some $F\colon Y \to \mathcal{P}(A) \setminus \{\emptyset\}$. At

$(\psi_1 \vee \psi_2, Y)$ Player 0 selects Y_1, Y_2 such that $Y_1 \cup Y_2 = Y$, and moves to $(\psi_i, Y_i \upharpoonright \mathit{free}(\psi_i))$ for an $i \in \{1,2\}$. Terminal positions are (ψ, Y), where Y is a literal. Finally, Win is defined as $\{(\psi, Y) : \psi \text{ is a literal}, \mathfrak{A} \models_Y \psi\}$. This model checking game for a formula evaluated with team semantics is essentially a model checking game for its translation into Σ^1_1. Note that the size of the game is exponentially larger than that of the first–order game for φ. Als, due to its involved definition it is unclear how to extract complexity results.

This motivates us to develop another apporach that is more revealing in the sense that it puts the complexity in the winning condition rather than in the game graph. In fact, this time we will keep the game graph as for FO. But the winning condition Win for Player 0 is now a subset of $\mathcal{P}(T)$ instead of a subset of T itself. To win, Player 0 needs a (non–deterministic) strategy $S = (W, F) \subseteq (V, E)$ such that the set of terminal positions reachable by a play from $X \subseteq I$ belongs to Win. In other words, if W is the set of nodes reachable from X by edges in F, then $W \cap T$ has to be in Win.

Theorem 4.5.1. The problem whether a given second–order reachability game $\mathcal{G}$ (with a compact description of Win) admits a winning strategy for Player 0 is NP–complete.

Proof. Clearly, we can efficiently guess a strategy S and check whether it is winning, hence the problem is in NP.

To show NP–hardness, we reduce SAT to it. Given a propositional formula φ in CNF consider $\mathcal{G}(\varphi)$, where

- Player 1 moves from φ to one of the clauses C,
- Player 0 moves from C to one of its literals (C, Y).

Terminal positions in the game are $T := \{(C, Y) : C \text{ clause of } \varphi,\ Y \in C\}$. The winning set is $\mathit{Win} := \{U \subseteq T \mid U \text{ contains no conflicting pair } (C, Y), (C', \neg Y)\}$. Then, 0 has a winning strategy from φ if, and only if, φ is satisfiable. □

Let us now define a model checking game for a logic with team semantics that extends FO by some of the several dependence atoms. Take the same game graph $\mathcal{G}(\mathfrak{A}, \varphi)$ as for FO and make sure that distinct occurrences of the same subformula are represented by distinct nodes. The set of initial positions is $I := \{(\varphi, s) : s \colon \mathit{free}(\varphi) \to A\}$. For any $U \subseteq V$ and ψ a subformula of φ, let $\mathit{Team}(U, \psi) := \{s : (\psi, s) \in U\}$. Then,

$$\mathit{Win} := \{U \subseteq T : \mathfrak{A} \models_{\mathit{Team}(U,\psi)} \psi \text{ for all literals } \psi\},$$

where $T := \{(\psi, s) \in V \mid \psi$ is a literal in $\varphi\}$.

By a straightforward induction, which we will not explicate here, the winning condition on $S = (W, F)$ translates from the literals to all subformulae, so that $S = (W, F)$ is a winning strategy from $X \subseteq I$ if, and only if, for all subformulae ψ of φ we have that $\mathfrak{A} \models_{Team(W,\psi)} \psi$.

Theorem 4.5.2. For every structure $\mathfrak{A}$ and every team X of assignments $s\colon free(\varphi) \to A$, we have that

$$\begin{aligned}\mathfrak{A} \models_X \varphi \Leftrightarrow\ & \text{Player 0 has a winning strategy for } \mathcal{G}(\mathfrak{A}, \varphi) \text{ from} \\ & I_X := \{(\varphi, s) : s \in X\},\ \text{i.e., a winning strategy } S = (W, F) \\ & \text{with } Team(W, \varphi) = X.\end{aligned}$$

Recall that in the previous theorem we showed that whether a given game graph $\mathcal{G}$ with a compact description for Win admits a winning strategy for Player 0 is NP–complete. Now, the size of a model checking game $\mathcal{G}(\mathfrak{A}, \varphi)$ on a finite structure $\mathfrak{A}$ is bounded by $|\varphi| \cdot |A|^{width(\varphi)}$, where $width(\varphi)$ is defined as $\max\{|free(\psi)| : \psi$ subformula of $\varphi\}$. Thus, the model checking problem for logics with team semantics on finite structure is in NEXPTIME and in NP for formula of bouned width (provided that atomic dependencies can be evaluated in polynomial time; the atoms we looked at certainly are). In fact, one can show that for dependence/independence/exclusion logic it is NEXPTIME–complete even for $\mathfrak{A} = \{0, 1\}$ and $X = \{\emptyset\}$, and NP–complete for formulae of bounded width.