A matter of trust: Dynamic attitudes in epistemic logic

Rodenhäuser, L.B.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5.
Logics

In this chapter, we move from a purely semantic to a properly logical setting, studying logical languages that contain operators for talking about dynamic attitudes in one way or another. We consider both the single-agent setting that we work with in most of this dissertation, and the multi-agent setting that we have introduced in §2.7.

The key construct of the single-agent language (studied in §§5.1–5.5) allows us to build sentences of the form

\[ [s: \varphi] \psi \]

with the reading “after the agent receives the information that \( \varphi \) from a source of type \( s \), \( \psi \) holds.” Here, a source’s being “of type \( s \)” is interpreted using the machinery developed in earlier chapters, i.e., we formalize the idea that to each type of source there corresponds a dynamic attitude capturing how the agent assesses the reliability of sources of that type.

The multi-agent language that we consider in this chapter (studied in §5.6) will allow us to build sentences of the form

\[ [a: \varphi] \psi \]

with the reading “after the communication act \([a: \varphi] \), \( \psi \) holds.” Here, the idea is that the agents use their dynamic attitude towards \( a \) to upgrade their plausibility order on an occasion where agent \( a \) asserts that \( \varphi \). This will be made precise using the work of §2.7, where we have introduced a formal notion of communication act.

The key notion we work with is the notion of a definable dynamic attitude. Very roughly and generally speaking, a dynamic attitude is said to be definable in a modal language if the language supplies syntactic devices that are expressive enough to describe the effects of applying certain upgrades to the models in which the language is interpreted. In this chapter, we consider the more specific case of definability in the epistemic-doxastic language.
which has operators for infallible and defeasible knowledge. The distinctive feature of the logics presented here compared to previous research is that they “work” not only for specific examples of belief revision policies, but for any dynamic attitude that is definable in the epistemic-doxastic language.

Introducing a logical setting with a proper syntax and semantics raises a number of familiar questions. We focus here on the two most basic, and most well-studied technical questions in dynamic epistemic logic, as they apply to our setting: expressivity and completeness.

5.1. The Epistemic-Doxastic Language

We have already seen the epistemic-doxastic language in §1.6. But since that has been a while ago, we develop all machinery from scratch.

5.1.1. Signatures. A signature is a tuple

\[(W, I, L, \Phi, A)\]

where

— \(W\) is a countably infinite set of possible worlds,
— \(I\) is a countable index set (called the set of attitude labels)
— \(L\) is a function assigning a dynamic attitude \(\tau\) to each attitude label \(\tau \in I\),
— \(\Phi\) is a set of symbols (called the set of atomic sentences), and
— \(A\) is a finite, non-empty set (called the set of agents)

The set \(A\) will only play a role for our discussion starting in §5.6, where we consider a multi-agent version of our setting. In the preceding sections, we will continue to discuss the single-agent setting familiar from most of the previous work in this dissertation, in which an agent receiving information, and the sources from which the agent receives this information, are merely implicit.

We shall refer, in this chapter, to attitude labels in \(I\) (which are to be thought of as “syntactic” objects) using lower-case greek letters (e.g., \(\sigma, \tau\)), and to the corresponding dynamic attitudes (semantic objects) assigned to the labels using the corresponding bold-face lower-case greek letters (e.g., \(\sigma, \tau\)), as we have already done in the preceding definition. That is, given some attitude label \(\tau\), we shall always assume that \(L(\tau)\) is given by \(\tau\).

Unless specifically noted otherwise, we assume an arbitrary but fixed signature \((W, I, L, \Phi, A)\) as given in the following.
5.1. The Epistemic-Doxastic Language

5.1.2. Valuations. A valuation $[\cdot]$ is a function

$$p \mapsto [p]$$

that assigns a proposition $P \subseteq W$ to each atomic sentence $p \in \Phi$.

Given a valuation $[\cdot]$, an atomic sentence $p \in \Phi$ and a proposition $P \subseteq W$, we shall write $[\cdot][p \rightarrow P]$ for the valuation which is just like $[\cdot]$, except that $[p] = P$.

5.1.3. Plausibility Models. A (single-agent) plausibility model is a pair

$$M = (S, [\cdot]),$$

where $S$ is a plausibility order (on $W$), and $[\cdot]$ is a valuation.

Given a plausibility model $M$, an atomic sentence $p \in \Phi$ and a proposition $P \subseteq W$, we shall write $M[p \rightarrow P]$ for the plausibility model $(S, [\cdot]^\prime)$ which is just like $M$ except that $[p] = P$.

5.1.4. The Epistemic-Doxastic Language. The language $L$ (called the (single-agent) epistemic-doxastic language) is given by the following grammar ($p \in \Phi$):

$$\varphi :: p \mid \neg \varphi \mid (\varphi \land \varphi) \mid \Box \varphi \mid K \varphi$$

Read $K \varphi$ as the agent infallibly (or: indefeasibly) knows that $\varphi$; read $\Box \varphi$ as the agent defeasibly knows that $\varphi$.

We define $\top$ as $p \lor \neg p$ and set $\bot := \neg \top$.

5.1.5. Semantics. We interpret the language $L$ in the usual manner, by providing, for each plausibility model $M = (S, [\cdot])$, a map $[\cdot]_M$ that assigns a proposition $[\varphi]_M \subseteq S$ to each sentence $\varphi \in \mathcal{L}$, the proposition comprising the worlds where $\varphi$ is satisfied in $M$ (or: the worlds $w$ such that $\varphi$ is true at $w$ in $M$).

Let $M = (S, [\cdot])$ be a plausibility model. $[\cdot]_M$ is defined by induction on the construction of $\varphi$.

$$[p]_M := [p] \cap S,$$
$$[-\varphi]_M := S \setminus [\varphi]_M,$$
$$[\varphi \land \psi]_M := [\varphi]_M \cap [\psi]_M,$$
$$[\Box \varphi]_M := \Box S [\varphi]_M,$$
$$[K \varphi]_M := KS [\varphi]_M.$$
where we recall the definitions of the propositional attitudes defeasible knowledge $\Box$ and irrecovable knowledge $K$ (cf. §1.2.8), according to which

$$\Box_S[\varphi]_M = \{ w \in S \mid \exists v \in S : v \leq_S w \text{ and } v \in [\varphi]_M \},$$

and

$$K_S[\varphi]_M = \{ w \in S \mid S \subseteq [\varphi]_M \}.$$  

We shall use the notation $M, w \models \varphi$ to mean that $w \in [\varphi]_M$. We say that a sentence $\varphi \in \mathcal{L}$ is valid iff $[\varphi]_M = S$ for any plausibility model $M = (S, \leq_S)$. We write $\models _\mathcal{L} \varphi$ if $\varphi$ is valid.

### 5.1.6. Axiomatization

The epistemic-doxastic language $\mathcal{L}$ was axiomatized by Baltag and Smets (2008) using the following derivation system. The logic of defeasible and indefeasible knowledge $\mathcal{L}$ is given by following axioms and rules of inference:

---

**Axioms:**

- All instances of theorems of propositional calculus
- $K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$
- $\Box (\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi)$
- The $S_5$ axioms for $K$
- The $S_4$ axioms for $\Box$
- $K\varphi \rightarrow \Box \varphi$
- $K(\varphi \lor \Box \psi) \land K(\psi \lor \Box \varphi) \rightarrow (K\varphi \lor K\psi)$

---

**Rules of inference:**

- From $\varphi$ and $\varphi \rightarrow \psi$ infer $\psi$
- From $\varphi$ infer $K\varphi$ and $\Box \varphi$

The fact that $\mathcal{L}$ is sound and complete w.r.t. plausibility models follows from the work of Baltag and Smets (2008). We argue as follows:

**Theorem 83 (Baltag and Smets (2008)).** The logic of defeasible and indefeasible knowledge $\mathcal{L}$ is weakly sound and complete w.r.t. plausibility models.

**Proof.** For soundness, observe that all the axioms are valid, and that the rules of inference preserve validity, and argue by induction on the length of a derivation in $\mathcal{L}$. For completeness, let $\varphi \in \mathcal{L}$ be an $\mathcal{L}$-consistent sentence (a sentence the negation of which is not provable in $\mathcal{L}$). Baltag and Smets (2008) show (their Theorem 2.5) that there exists a triple $\mathcal{X} = (X, \leq, V)$, where $X$ is a finite, non-empty set, $\leq$ a total preorder on $X$ and $V$ a function assigning a set
of worlds $Y \subseteq X$ to each atomic sentence $p \in \Phi$, and an element $x \in X$ such that $\mathcal{X}, x \models_{BS} \varphi$, with $\models_{BS}$ their truth predicate.

Given some such $\mathcal{X} = (X, \leq_X, V)$ and $x \in X$ such that $\mathcal{X}, x \models_{BS} \varphi$, we choose a set $S \subseteq W$ with the same cardinality as $X$ (an appropriate $S$ can be found since $X$ is finite, while $W$ is countably infinite). Pick any bijection $\iota : W \rightarrow X$, and put:

- for any $w, v \in S$: $w \leq_S v$ iff $\iota(w) \leq_X \iota(v)$,
- for any $p \in \Phi$: $\llbracket p \rrbracket := \{ w \in S \mid \iota(w) \in V(p) \}$.

By definition, $M$ and $\mathcal{X}$ are isomorphic along the bijection $\iota$. Furthermore, the satisfaction relation $\models_{BM}$ defined in [Baltag and Smets 2008] runs exactly as the definition of our satisfaction relation $\models_{\mathcal{X}}$. So $\mathcal{X}, x \models_{BS} \varphi$ iff $M, \iota^{-1}(x) \models \varphi$. Thus, by the fact that $\mathcal{X}, x \models_{BS} \varphi$, it follows that $M, \iota^{-1}(x) \models \varphi$.

This argument shows that every $L$-consistent sentence $\varphi \in L$ is satisfiable in a non-empty plausibility model, and from this observation, completeness follows in the usual manner.

5.2. Definable Dynamic Attitudes

5.2.1. Upgrades on Plausibility Models. We have introduced upgrades as transformations of plausibility orders (cf. §1.3). For use in the following, we would like to think of upgrades as applying to plausibility models. Since our setting only models epistemic changes, i.e., changes in the information state of some agent, and not changes of the basic (“ontic”, agent-independent) facts of the world, this turns out to be simply a matter of introducing appropriate notation.

Namely, given a plausibility model $M = (S, \llbracket \cdot \rrbracket)$, a dynamic attitude $\tau$, and a proposition $P$, we put

$$M^{\tau P} := (S^{\tau P}, \llbracket \cdot \rrbracket).$$

So applying the upgrade $\tau P$ to the plausibility model $M$ amounts to applying $\tau P$ to $S$, and simply “dragging along” the valuation $\llbracket \cdot \rrbracket$: the propositions to which atomic sentences evaluate stay the same when applying an upgrade.

5.2.2. Definable Dynamic Attitudes. For the remainder of this chapter, fix two distinct atomic sentences $p, q \in \Phi$, and let $\tau$ be a dynamic attitude.

---

1In other words, our semantics is a notational variant of theirs, with the single difference that the semantics of [Baltag and Smets 2008] also allows infinite structures.
A sentence $\vartheta \in \mathcal{L}$ (possibly containing occurrences of $p_*$ and $q_*$) defines $\tau$ (in $\mathcal{L}$) iff for any plausibility model $\mathcal{M} = (S, [\_])$ and world $w \in S$:

$$\mathcal{M}, w \models \vartheta \iff \mathcal{M}^{[p_*]}, w \models \Diamond q_*.$$ 

The displayed equivalence expresses that $\vartheta$ captures precisely the set of worlds satisfying $\Diamond q_*$ in the plausibility model $\mathcal{M}^{[p_*]}$ resulting from upgrading $\mathcal{M}$ with $\tau^{[p_*]}$. Since we require this to be the case in all plausibility models, $\vartheta$ can be said to pre-encode the effect of learning $[p_*]$ from a $\tau$-source as far as the defeasible possibility of $[q_*]$ is concerned.

We will show in §5.4 that this is actually tantamount to saying (in a sense to be made precise) that adding operators encoding the dynamic attitude $\tau$ to our language $\mathcal{L}$ does not actually add expressive to $\mathcal{L}$, and this will be crucial to prove completeness for our languages that allow us to talk about dynamic attitudes.

If a sentence $\vartheta$ defines a dynamic attitude $\tau$, then we call $\vartheta$ a definition of $\tau$ (in $\mathcal{L}$). A dynamic attitude $\tau$ is called definable (in $\mathcal{L}$) if there exists a definition of $\tau$.

### 5.2.3. Examples of Definitions

Here are some examples of definitions of dynamic attitudes.

#### Proposition 84.

1. $p_* \land \Diamond (p_* \land q_*)$ defines infallible trust $!$.
2. $\Diamond (p_* \land q_*) \lor (\neg p_* \land (\Diamond (\neg p_* \land q_*) \lor K^-(p_* \land q_*)))$ defines strong trust $\dashv$.
3. $\Diamond q_*$, defines neutrality $\text{id}$.
4. $\bot$ defines isolation $\emptyset$.

#### Proof.

1. Let $\mathcal{M} = (S, [\_])$ be a plausibility model, and let $w \in S$. We have to show that $\mathcal{M}, w \models p_* \land \Diamond (p_* \land q_*)$ iff $\mathcal{M}^{[p_*]}, w \models \Diamond q_*$. Observe that $\mathcal{M}, w \models p_* \land \Diamond (p_* \land q_*)$ iff (by the semantics) $w \in [p_*]$ and there exists $v \in S$: $v \leq_S w$ and $v \in [p_*] \cap [q_*]$ iff (by definition of $!$) $w \in S^{\tau^{[p_*]}}$ and there exists $v \in S^{\tau^{[p_*]}}$: $v \leq_{S^{\tau^{[p_*]}}} w$ and $v \in [q_*]$ iff (by the semantics) $\mathcal{M}^{\tau^{[p_*]}}, w \models \Diamond q_*$. This shows our claim, so $p_* \land \Diamond (p_* \land q_*)$ defines infallible trust $!$.

2. Let $\mathcal{M} = (S, [\_])$ be a plausibility model, and let $w \in S$. We have to show that $\mathcal{M}, w \models \Diamond (p_* \land q_*) \lor (\neg p_* \land (\Diamond (\neg p_* \land q_*) \lor K^-(p_* \land q_*)))$ iff $\mathcal{M}^{\tau^{[p_*]}} \models \Diamond q_*$. 


5.2. Definable Dynamic Attitudes

From left to right, suppose that
\[ M, w \models \diamond (p_\ast \land q_\ast) \lor (\neg p \land (\diamond (\neg p_\ast \land q_\ast) \lor K^{-1} (p_\ast \land q_\ast))). \]

This implies that (1.) or (2.) below holds:

1. \[ M, w \models \diamond (p_\ast \land q_\ast) \]
2. \[ M, w \models (\neg p_\ast \land (\diamond (\neg p_\ast \land q_\ast) \lor K^{-1} (p_\ast \land q_\ast))). \]

If (1) holds, then there exists \( v \) such that \( v \leq_S w \) and \( v \in [p_\ast] \cap [q_\ast] \). By definition of \( \llbracket \cdot \rrbracket \), \( v \llbracket p_\ast \rrbracket \) \( w \), hence \( M \llbracket p_\ast \rrbracket, w \models \diamond q_\ast \). If (2) holds, then either \( M, w \models \neg p_\ast \land (\diamond (\neg p_\ast \land q_\ast) \lor K^{-1} (p_\ast \land q_\ast)) \) or (2a) \( M, w \models (\neg p_\ast \land q_\ast). \) If (2a) holds, then \( M, w \not\models p_\ast \) and there exists \( v \) such that \( v \leq_S w \) and \( M, v \not\models p_\ast \), and \( M, v \models q_\ast \). By definition of \( \llbracket \cdot \rrbracket \), \( v \llbracket p_\ast \rrbracket \) \( w \), hence \( M \llbracket p_\ast \rrbracket, w \models \diamond q_\ast \). If (2b) holds, then \( M, w \not\models p_\ast \) and there exists \( v \in S \) such that \( M, v \models p_\ast \land q_\ast \). By definition of \( \llbracket \cdot \rrbracket \), \( v \llbracket p_\ast \rrbracket \) \( w \). Hence \( M, w \models \diamond q_\ast \). So in either of the two cases (1) or (2), our claim holds, and this concludes the left to right direction.

From right to left, suppose that \( M \llbracket p_\ast \rrbracket, w \models \diamond q_\ast \). Thus exists \( v \in S \) such that \( v \leq_S w \) and \( v \in [q_\ast] \). We argue in two cases. First, suppose that \( w \in [p_\ast] \). Then also \( v \in [p_\ast] \) by definition of \( \llbracket \cdot \rrbracket \). Furthermore, again by definition of \( \llbracket \cdot \rrbracket \), it is the case that \( v \leq_S w \). Since \( v \in [q_\ast] \), we conclude that \( M, w \models \diamond (p_\ast \land q_\ast) \). So our claim holds in the first case. Second, suppose that \( w \not\in [p_\ast] \). Distinguish two sub-cases. First, suppose that \( v \not\in [p_\ast] \). Then \( v \leq_S w \) and since \( v \in [q_\ast] \), it follows that \( M, w \models \neg p_\ast \land (\diamond (\neg p_\ast \land q_\ast) \lor K^{-1} (p_\ast \land q_\ast)). \) So our claim holds in the first sub-case. For the second sub-case, suppose that \( v \in [p_\ast] \). Then \( v \leq_S w \) and since \( v \in [q_\ast] \), it follows that \( M, w \models (\neg p_\ast \land K^{-1} (p_\ast \land q_\ast)). \) So our claim holds in the second sub-case. Thus our claim holds in the second case. In either case, our claim holds, and this concludes the direction from right to left.

3. Let \( M = (S, [\cdot]) \) be a plausibility model, and let \( w \in S \). We have to show that \( M, w \models \diamond q_\ast \) iff \( M^{id}[p_\ast], w \models \diamond q_\ast \). Since \( M^{id}[p_\ast] = M \), this is indeed the case.

4. Let \( M = (S, [\cdot]) \) be a plausibility model, and let \( w \in S \). We have to show that \( M, w \models \perp \) iff \( M^{\Theta}[p_\ast], w \models \diamond q_\ast \). Since neither \( M, w \models \perp \) nor \( M^{\Theta}[p_\ast], w \models \diamond q_\ast \), the claim holds.

5.2.4. Non-Definable Attitudes. An example of a dynamic attitude that is, regrettably, \underline{not} definable in \( L \), is minimal trust \( \uparrow \). To show this, we introduce an appropriate notion of bisimulation for our epistemic-doxastic language \( L \) (Blackburn, de Rijke, and Venema 2001).
Given a plausibility order $S$, we write $w \sim_S v$ iff $w, v \in S$, and we write $w \geq_S v$ iff $v \leq_S w$. Now let $M = (S, \llbracket \cdot \rrbracket)$ and $M' = (S', \llbracket \cdot \rrbracket')$ be plausibility models. A bisimulation between $M$ and $M'$ is a relation $R \subseteq S \times S'$ such that for any $w \in S$, $w' \in S'$: if $w R w'$, then

- for any $p \in \Phi$: $w \in [p]$, iff $w' \in [p]'$,
- if there exists $v$: $w \sim_S v$, then there exists $v'$: $w' \sim_{S'} v'$ and $v R v'$, and vice versa,
- if there exists $v$: $w \geq_S v$, then there exists $v'$: $w' \geq_{S'} v'$ and $v R v'$, and vice versa.

We write $M, w \models M', w'$ iff there exists a bisimulation $R$ between $M$ and $M'$ such that $w R w'$.

**Proposition 85.** Suppose that $M, w \models M', w'$. Then for any $\varphi \in L$:

$$M, w \models \varphi \text{ iff } M', w' \models \varphi.$$

**Proof.** Routine (cf. Blackburn et al. (2001)).

We apply the previous proposition to show that $\uparrow$ is not definable. The proof uses Figure 24: the dotted lines in the Figure indicate a bisimulation between the two plausibility models $M$ (to the left) and $M'$ (to the right). Notice that $M'[p], z \models \Diamond q$, while $(M')'[p]'$, $v \not\models \Diamond q$. However $M, z \models M', v$, so the two model-world pairs agree, by the above proposition, on all sentences in $L$. Given these facts, a glance at the definition of definability in §5.2.2 is enough to convince us that $\uparrow$ is not definable. We repeat the argument more formally in the next proposition.

**Proposition 86.** Minimal trust $\uparrow$ is not definable in $L$.

**Proof.** Consider the two plausibility models $M = (S, \llbracket \cdot \rrbracket)$ (depicted on the left of Figure 24) and $M' = (S', \llbracket \cdot \rrbracket')$ (depicted on the right of Figure 24). Clearly, $M, z \models M', v$. Now let $\theta \in \mathcal{L}$ and suppose that $\theta \in \mathcal{L}$ defines $\uparrow$ in $\mathcal{L}$. Observe that $M, z \models \theta$, since $M[\llbracket p \rrbracket], z \models \Diamond q_*$. From the fact that $M, z \models M', v$, it follows by the previous proposition that $M', v \models \theta$. Since $\theta$ defines $\tau$, we conclude that $(M')[\llbracket p \rrbracket'], v \models \Diamond q_*$. But the latter statement is false. Hence $\theta$ does not define $\uparrow$ in $\mathcal{L}$. So $\uparrow$ is not definable in $\mathcal{L}$.

As a consequence, the setting we consider in this chapter can define only a selection of the important examples of dynamic attitudes considered in this dissertation. The first question this raises is what definability in $L$ exactly amounts to: just which dynamic attitudes are definable in $L$? Second, the
result points to the fact that the techniques presented in this chapter should ultimately be adapted to more expressive languages. One promising candidate is the language which has the modality $K$, and, in addition, modalities for the strict plausibility order $<_S$, and for the equiplausibility order $\sim_S$ (in a given plausibility order $S$). As observed in Baltag and Smets (2008), this language is strictly more expressive than the epistemic-doxastic language $L$ considered here.

Third, weaker languages might also be considered. For example, suppose we add conditional belief operators $B_{\psi} \phi$ to propositional logic (interpreted in the natural way, using our definition of the propositional attitude $B_{\phi} P$ in §1.2.7). The results of van Benthem (2007) essentially show that this language can define minimal trust $\hat{\psi}$. The counterexample exposed in Figure 24 does not apply, since the two bisimilar worlds $z$ and $v$ also agree on all sentences in the language with conditional belief operators after applying an upgrade $\uparrow P$. Interestingly, then, it is not always the case that a strictly more expressive language can define strictly more dynamic attitudes! The precise formulation and study of a more general notion of definability, applicable to the languages mentioned, is a topic for future research.

\footnote{In particular, notice that $\Box \phi$ is equivalent to $\neg \phi \land [\prec] \phi$.}
5.3. Languages for Definable Sources

In this section, we construct logical languages for (implicit) agents that receive information from sources that are “definable” in the sense that the dynamic attitude of the agents towards her sources is given by definable dynamic attitudes in the sense of the previous section. We shall assume that each agent has a finite number of such sources at her disposal. Since the dynamic attitude of the agent towards her sources are definable, we can collect suitable definitions for these dynamic attitudes. This leads to the notion of a source structure, which will be an important parameter for our logics.

5.3.1. Definable Sources. A source structure is a pair

\[ \Sigma = (\text{Att}, \text{Def}) \]

such that \( \text{Att} \subseteq I \) is a finite, non-empty set of attitude labels and \( \text{Def} : \text{Att} \rightarrow \mathcal{L} \) assigns a sentence \( \text{Def}(\tau) \in \mathcal{L} \) to each \( \tau \in \text{Att} \), in such a way that \( \text{Def}(\tau) \) defines \( \tau \).

Given a source structure \( \Sigma = (\text{Att}, \text{Def}) \), a (definable) source (over \( \Sigma \)) is a pair

\[ s := (\tau_s, \vartheta_s), \]

where \( \tau_s \in \text{Att} \), and \( \vartheta_s = \text{Def}(\tau_s) \). By abuse of notation, we write \( s \in \Sigma \) iff \( s \) is a source over \( \Sigma \).

We now use source structures as an additional parameter to construct languages extending our epistemic-doxastic language \( \mathcal{L} \).

5.3.2. The extended language \( \mathcal{L}[\Sigma] \). For each source structure \( \Sigma \), we obtain the language \( \mathcal{L}[\Sigma] \) by adding a new construction rule to the grammar for \( \mathcal{L} \), allowing us to form sentences of the form \([s: \varphi] \psi\), with \( s \in \Sigma \).

Formally, \( \mathcal{L}[\Sigma] \) is given by the following grammar:

\[ \varphi ::= p \mid \neg \varphi \mid (\varphi \land \varphi) \mid \Box \varphi \mid K \varphi \mid [s: \varphi] \psi, \]

where \( p \in \Phi \), and \( s \in \Sigma \).

Read \([s: \varphi] \psi\) as “after the agent receives the information that \( \varphi \) from a source of type \( s \), \( \psi \) holds.”

---

3Recall that the index set \( I \) is part of the fixed signature we assume as given, cf. \( \S 5.1.1 \).
5.3.3. **Semantics of \( \mathcal{L}[\Sigma] \).** To obtain a semantics for \( \mathcal{L}[\Sigma] \), we extend the semantics for \( \mathcal{L} \) accordingly. We define, given a plausibility model \( \mathcal{M} \):

\[
\begin{align*}
[p]_{\mathcal{M}} & := [p] \cap S, \\
[\neg \varphi]_{\mathcal{M}} & := S \setminus [\varphi]_{\mathcal{M}}, \\
[\varphi \land \psi]_{\mathcal{M}} & := [\varphi]_{\mathcal{M}} \cap [\psi]_{\mathcal{M}}, \\
[\Box \varphi]_{\mathcal{M}} & := \Box S [\varphi]_{\mathcal{M}}, \\
[K \varphi]_{\mathcal{M}} & := KS[\varphi]_{\mathcal{M}}, \\
[[s: \varphi] \psi] & := S \cap (S^{T_s}[\varphi]_{\mathcal{M}} \Rightarrow [\varphi]_{\mathcal{M}}^{T_s}[\psi]_{\mathcal{M}}).
\end{align*}
\]

Notice that, except for the last clause, this just repeats the definition of the semantics for \( \mathcal{L} \) given in §5.1.4. Notice furthermore that, in the last clause, for any world \( w \in S \), we check whether \( w \) satisfies the proposition \([\varphi]_{\mathcal{M}}^{T_s}[\psi]_{\mathcal{M}}\), provided \( w \) is contained in the domain of the upgrade model \( \mathcal{M}^{T_s}[\varphi]_{\mathcal{M}}\). As usual in many dynamic epistemic logics, \([[[s: \varphi] \psi]]_{\mathcal{M}}\) contains in particular those worlds in \( S \) which are not contained in \( S^{T_s}[\varphi]_{\mathcal{M}}\): these worlds trivially satisfy the sentence \([s: \varphi] \psi\).

As before, we write \( \mathcal{M}, w \models \varphi \) to mean that \( w \in [\varphi]_{\mathcal{M}} \). We say that a sentence \( \varphi \in \mathcal{L}[\Sigma] \) is valid iff \( [\varphi]_{\mathcal{M}} = S \) for any plausibility model \( \mathcal{M} = (S, [\ ]\ ) \). We write \( \models_{\mathcal{L}[\Sigma]} \varphi \) if \( \varphi \) is valid.

5.3.4. **Reintroducing Dynamic Attitudes in the Notation.** Given a source structure \( \Sigma \), a source \( s = (\tau_s, \vartheta_s) \in \Sigma \), and sentences \( \varphi, \psi \in \mathcal{L}[\Sigma] \), we put

\[
[\tau_s \varphi] \psi := [[s: \varphi] \psi].
\]

This yields notation familiar from the existing literature (van Benthem 2007, Baltag and Smets 2008), like \([! \varphi] \psi\), \([\uparrow \varphi] \psi\) (assuming a suitable signature and source structure). But note that, in our setting, \([! \varphi] \psi\) and \([\uparrow \varphi] \psi\) are not official syntax, but defined notation.

### 5.4. Expressivity

This section is devoted to showing that for our languages \( \mathcal{L}[\Sigma] \), parametrized by a choice of source structure \( \Sigma \), we can obtain “reduction axioms” in the usual style of dynamic epistemic logic. As a consequence, adding definable sources to \( \mathcal{L} \) yields no increase in expressive power to \( \mathcal{L} \).

\[^{4}\text{Cf. van Ditmarsch, Kooi, and van der Hoek (2007).}\]
5.4.1. Notation. Let $\Sigma$ be a source structure, let $s \in \Sigma$, let $\varphi, \psi \in \mathcal{L}[\Sigma]$ be sentences. We write $\varphi[p \rightarrow \psi]$ for the sentence resulting from $\varphi$ by simultaneously substituting all occurrences of $p$ in $\varphi$ with $\psi$.

The following notation will be useful:

- $\diamondsuit s(\varphi, \psi) := \vartheta_s[p \rightarrow \varphi, q \rightarrow \psi]$,
- $\Box s(\varphi, \psi) := \neg \diamondsuit s(\varphi, \psi)$,
- $\text{pre}_s(\varphi) := \diamondsuit s(\varphi, \top)$.

Recalling that $s := (\tau_s, \vartheta_s)$, with $\tau_s$ an attitude label, and $\vartheta_s$ a definition of $\tau_s$ given by the source structure, we notice that $\diamondsuit s(\varphi, \psi)$ is the definition of $\tau_s$ (given by $\Sigma$) with $p_*$ substituted by $\varphi$ and $q_*$ substituted by $\psi$.

The notation suggests that our definitions of dynamic attitudes actually serve the purpose of behaving very much like binary modal operators, and this is indeed just what we want to establish in the following.

It is then also natural to define a “dual” $\Box s(\varphi, \psi)$ of $\diamondsuit s(\varphi, \psi)$, which is just what we do in the second notation introduced above.

Finally, the third piece of notation $\text{pre}_s(\varphi)$ is meant to suggest that using our definitions of dynamic attitudes, we can capture “pre-conditions” (conditions of executability) for applying upgrades, i.e., we would like the sentence $\text{pre}_s(\varphi)$ to capture the worlds (in some given plausibility order $S$) where the information that $\varphi$ can be received from the source $s$.

The remainder of this section is devoted to putting the above notation to work, and making the above informal remarks precise. We start with a simple example.

5.4.2. Example. Using our new notation, we can “rewrite” the usual reduction axioms familiar from dynamic epistemic logic in a generic format. Consider $p_* \land \diamondsuit(p_* \land q_*)$, which defines, as we have seen above, the dynamic attitude $!$. Let $\Sigma$ be a source structure, and let $s \in \Sigma$ be a source such that $\vartheta_s = p_* \land \diamondsuit(p_* \land \top)$ (i.e., $\vartheta_s$ defines $!$). Our starting point is the observation that

$$[!p_*]q_* \leftrightarrow (p_* \rightarrow q_*)$$

is a valid sentence of $\mathcal{L}[\Sigma]$. In this equivalence, $p_*$ on the right hand side functions as the precondition of $!p_*$. We can now “rewrite” the above equivalence using our notation $\text{pre}_s(\varphi)$. Notice that $\text{pre}_s(p_*)$ works out to $p_* \land \diamondsuit(p_* \land \top)$, which is equivalent to $p_*$, true exactly at the worlds contained in $S^{|!p_*|}$. So

$$[!p_*]q_* \leftrightarrow (\text{pre}_s(p_*) \rightarrow q_*)$$

is also valid. Now consider a sentence $[!p_*]\Box q_*$. Observe that

$$[!p_*]\Box q_* \leftrightarrow (p_* \rightarrow \Box[!p_*]q_*)$$
is valid. This is equivalent to

\[ ![p] \Box q, \leftrightarrow (p, \rightarrow (p, \rightarrow [![p] \Box q))). \]

As we have seen above, \( p, \) is equivalent to \( \text{pre}_s(p,). \) Furthermore, we notice that \( p, \rightarrow [![p] \Box q), \) is equivalent to \( \Box_s(p, [![p] \Box q). \) Substituting equivalents, we obtain that

\[ ![p] \Box q, \leftrightarrow (\text{pre}_s(p,) \rightarrow \Box_s(p, [![p] \Box q)) \]

is valid. So we can write a reduction law encoding the dynamics of the defeasible knowledge operator \( \Box \) under applying updates using the formal machinery introduced so far. While, in the above example, we have only worked with atomic sentences, we now consider laws of this kind more generally.

### 5.4.3. Reduction Laws

As a matter of general fact, source structures give rise to languages in which “reduction laws” in the usual style of dynamic epistemic logic are valid.

We first observe that, given a source structure \( \Sigma, \) we can write down preconditions (conditions of executability) for each source \( s \in \Sigma. \)

**Lemma 87.** Let \( \Sigma, \) be a source structure, and \( s, \in \Sigma. \) Then for any sentence \( \varphi \in \mathcal{L}[\Sigma] \) and plausibility model \( M \): \( [\text{pre}_s(\varphi)]_M = S^\tau_s[\varphi]|_M. \)

**Proof.** Let \( \Sigma, \) be a source structure, and \( s, \in \Sigma. \) Let \( \varphi \in \mathcal{L}[\Sigma] \) and let \( M = (S,[\_]) \) be a plausibility model. We have to show that \( [\text{pre}_s(\varphi)]_M = S^\tau_s[\varphi]|_M. \) Recall that \( \text{pre}_s(\varphi) = \vartheta_s([\varphi],[\_]). \) Let \( [\_]' := [\_][p, \rightarrow [\varphi]|_M[q, \rightarrow W], \) and let \( N = (S,[\_]'). \) By definition of \( N, \) we have \( [\text{pre}_s(\varphi)]_M = [\vartheta_s]|_N. \) Since \( \vartheta_s \) defines \( \tau_s, \) we have \( [\vartheta_s]|_N = [\Diamond q]|_N^{\tau_s[p]}|_N. \) By definition of \( [\_]' : [\vartheta_s]|_N = [\Diamond \top]|_N^{\tau_s[p]}. \) By the semantics: \( [\Diamond \top]|_N^{\tau_s[p]} = S^\tau_s[p]|_N. \) By definition of \( [\_]' : S^\tau_s[p]|_N = S^\tau_s[\varphi]|_M. \) Overall, we have established that \( [\text{pre}_s(\varphi)]_M = S^\tau_s[\varphi]|_M, \) the desired result.

According to the preceding lemma, given a source \( s, \in \Sigma \) and a plausibility model \( M = (S,[\_]), \) the sentence \( \text{pre}_s(\varphi) \) is satisfied in just those worlds \( w, \in S \) in which the upgrade \( \tau_s[\varphi]|_M \) is executable (cf. §3.1.3.2 for the definition of executability). In this sense, \( \text{pre}_s(\varphi) \) captures the precondition of \( \tau_s[\varphi]|_M. \)

Next, we observe, for use in the proof of Proposition 88 below:

**Lemma 88.** For any source structure \( \Sigma, \) source \( s, \in \Sigma, \) sentences \( \varphi, \psi \in \mathcal{L}[\Sigma], \) plausibility model \( M = (S,[\_]), \) and world \( w, \in S^\tau_s[\varphi]|_M\):

\[ M,w \models [s: \varphi] \Box \psi \quad \text{iff} \quad M,w \models \Box_s(\varphi, [s: \varphi]|_\psi). \]
Chapter 5. Logics

Proof. Let $\Sigma$ be a source structure, $s \in \Sigma, \varphi, \psi \in L[\Sigma], M = (S, [\cdot])$ a plausibility model, and $w \in S^{\tau_s}[\psi]_M$.

We define the plausibility model $N$ as

$$N := M[p \mapsto [\varphi]_M, q \mapsto [[s: \varphi]_M]_M].$$

Now coming from the left,

$$M, w = [s: \varphi] \Box \psi$$

iff (by the semantics and the assumption that $w \in S^{\tau_s}[\psi]_M$)

$$M^{\tau_s}[\varphi]_M, w \models \Box \psi$$

iff (by the semantics)

$$S^{\tau_s}[\varphi]_M, w \models \Box [\psi]_M^{\tau_s}[\varphi]_M.$$

On the other hand, coming from the right,

$$M, w \models \Box_s (\varphi, [s: \varphi] \psi))$$

iff (by definition of $N$)

$$N, w \models \Box_s (p, q)$$

iff (since $\varphi_s$ defines $\tau_s$ and by the assumption that $w \in S^{\tau_s}[\varphi]_M$)

$$N^{\tau_s}[\varphi]_N, w \models \Box q$$

iff (by the definition of $N$)

$$N^{\tau_s}[\varphi]_M, w \models \Box q$$

iff (by the semantics)

$$S^{\tau_s}[\varphi]_M, w \models [q]_N.$$

Letting $S' := S^{\tau_s}[\varphi]_M$, to prove our desired equivalence it is thus sufficient to show that $\Box S'[\varphi]_M^{\tau_s}[\varphi]_M = \Box S'[q]_N$. By the semantics, $\Box S'[q]_N = \Box S'(S' \cap [q]_N)$. Noticing that, by definition of $N$, we have that $[q]_N = [[s: \varphi]_M]_N$, it follows using the semantics that $\Box S'[q]_N = \Box S'(S' \cap (S' \Rightarrow [\psi]_M^{\tau_s}[\varphi]_M))$, hence $\Box S'[q]_N = \Box S'[\varphi]_M^{\tau_s}[\varphi]_M$, the desired result.

The preceding lemma is the heart of the proof, in Proposition 89 below, that establishes the validity of the reduction law for defeasible knowledge $\Box$.

Proposition 89. For any source structure $\Sigma$, the following are valid in $L[\Sigma]$:

1. $[s: \varphi] p \leftrightarrow (\text{pre}_s(\varphi) \rightarrow p)$,
5.4. Expressivity

— \([s: \varphi] \psi \leftrightarrow (\text{pre}_s(\varphi) \rightarrow \neg[s: \varphi] \psi)\),
— \([s: \varphi](\varphi \land \chi) \leftrightarrow ([s: \varphi]\psi \land [s: \varphi]\chi),
— \([s: \varphi]K\psi \leftrightarrow (\text{pre}_s(\varphi) \rightarrow K[s: \varphi]\psi),
— \([s: \varphi]\Box \psi \leftrightarrow (\text{pre}_s(\varphi) \rightarrow \Box_s(\varphi,[s: \varphi]\psi)).

**Proof.** We consider the most interesting item, the last one. Let \(\Sigma\) be a source structure, let \(s \in \Sigma\), let \(M = (S, [\cdot])\) be a plausibility model, and let \(w \in S\).

We notice, using Lemma \[87\] that if \(\tau_s[\varphi]_M\) is not executable in \(w\), then both sides of the bi-implication in (5.) are satisfied at \(w\), i.e., under this assumption, \(M, w \models [s: \varphi] \Box \psi\) and \(M, w \models (\text{pre}_s(\varphi) \rightarrow \Box_s(\varphi,[s: \varphi]\psi))\) hold trivially, so our claim holds as well. For the remainder of the proof, we may thus assume that \(\tau_s[\varphi]_M\) is executable in \(w\). Suppose, then, that \(w \in S^{\tau_s[\varphi]}_M\), which, by Lemma \[87\] is equivalent to saying that \(M, w \models \text{pre}_s(\varphi)\). Under this assumption, we observe that

\[M, w \models \text{pre}_s(\varphi) \rightarrow \Box_s(\varphi,[s: \varphi]\psi) \iff M, w \models \Box_s(\varphi,[s: \varphi]\psi).\]

But the fact that

\[M, w \models [s: \varphi] \Box \psi \iff M, w \models \Box_s(\varphi,[s: \varphi]\psi).\]

is exactly Lemma \[88\] and we are done.

5.4.4. Complexity. Our aim is to show that every sentence \(\varphi \in \mathcal{L}^{S}\) is semantically equivalent to a sentence \(\psi \in \mathcal{L}\) in the sense that \(\models_{\mathcal{L}[\Sigma]} \varphi \leftrightarrow \psi\), using the above reduction laws.\[3\] The proof hinges on an appropriate measure of (syntactic) complexity for sentences of \(\mathcal{L}^{S}\), which we proceed to define.

For each source structure \(\Sigma\), for each sentence \(\varphi \in \mathcal{L}^{S}\), the *horizontal depth* \(h(\varphi)\) of \(\varphi\) is inductively given by

\[
\begin{align*}
    h(p) &:= 0, \\
    h(\neg \varphi) &:= h(\varphi), \\
    h(\varphi \land \psi) &:= \max\{h(\varphi), h(\psi)\}, \\
    h(K\varphi) &:= h(\varphi), \\
    h(\Box \varphi) &:= h(\varphi), \\
    h([s: \varphi] \psi) &:= 1 + h(\varphi).
\end{align*}
\]

The horizontal depth measures the extent to which dynamic operators are (“horizontally”) stacked. For example, \(\varphi_1 = [s: \varphi][s: \psi] \chi\) is more horizontally complex than \(\varphi_2 = [s: \varphi] \psi\) in the sense that \(h(\varphi_1) > h(\varphi_2)\). Intuitively, one

\[\text{Our strategy for establishing this follows Kooi (2007).}\]
scans a given sentence from left to right, counting dynamic operators in the
scope of others.

For each source structure $\Sigma$, for each sentence $\phi \in L[\Sigma]$, the \textit{vertical depth}
$v(\phi)$ of $\phi$ is inductively given by

- $v(p) := 0$
- $v(\neg \phi) := v(\phi)$
- $v(\phi \land \psi) := \max\{v(\phi), v(\psi)\}$
- $v(\boxdot \phi) := v(\phi)$
- $v([s \cdot \phi] \psi) := \max\{v(\phi) + 1, v(\psi)\}$

The vertical depth measures the extent to which dynamic operators occur
(“vertically”) nested inside of other dynamic operators (e.g., $\psi_1 = [s : [s : \phi] \psi] \chi$
is more vertically complex than $\psi_2 = [s : \phi] \chi$ in the sense that $v(\psi_1) > v(\psi_2)$).
Intuitively, one jumps “inside” the dynamic operators, counting the number
of jumps needed until one hits “the bottom of the box.”

For each source structure $\Sigma$, for each sentence $\phi \in L[\Sigma]$, the \textit{complexity} $c(\phi)$
of $\phi$ is given by

$$c(\phi) := h(\phi) \cdot v(\phi).$$

So the complexity of a sentence is obtained by multiplying its horizontal
and its vertical depth.

We observe that sentences in $L$ have complexity 0. Moreover, if a sen-
tence in $L[\Sigma]$ has complexity 0, then it must be a sentence in $L$. In a concise statement:

**Lemma 90.** For every source structure $\Sigma$, for every sentence $\phi \in L[\Sigma]$: $\phi \in L$ iff
$c(\phi) = 0$.

**Proof.** A trivial induction on $\phi \in L$ shows the left to right direction. For the
other direction, suppose that $\phi \in L[\Sigma]$, $\phi \notin L$. Then $\phi$ contains a subformula
of the form $[s : \psi] \chi$. By definition of $c$, it follows that $c(\phi) \geq 1$, and this shows
the other direction.

---

5.4.5. \textbf{Reduction of $L[\Sigma]$ to $L$.} Intuitively, our reduction laws give us the
tools to reduce the complexity $c(\phi)$ of a given sentence $\phi$ in $L[\Sigma]$ until we
eventually obtain a sentence $\psi$ of complexity 0, with $\psi$ equivalent to our original $\phi$; by the previous lemma, $\psi$ will be a sentence in $L$. Let us put this idea
into practice.

For $x \in \{v, h, c\}$, we write $\phi \leq_x \psi$ iff $x(\phi) \leq x(\psi)$; $\phi <_x \psi$ iff $x(\phi) < x(\psi)$; and $\phi =_x \psi$ iff $x(\phi) = x(\psi)$.
5.4. Expressivity

We first show that, assuming that we already have a sentence \( \varphi \in \mathcal{L} \), we can always properly reduce the vertical depth of the sentence \([s:\varphi]\varphi\), for any \( \varphi \in \mathcal{L}[\Sigma] \), getting rid of the dynamic modality \([s:\varphi]\), using the reduction laws.

**Lemma 91.** For every source structure \( \Sigma \), for every sentence \( \varphi \in \mathcal{L} \), for every sentence \( \psi \in \mathcal{L}[\Sigma] \), there exists a sentence \( \chi \in \mathcal{L}[\Sigma] \) such that \( \chi \equiv_{\psi} [s: \psi] \varphi \) and \( \models_{\Sigma} \chi \leftrightarrow [s: \psi] \varphi \).

**Proof.** Let \( \Sigma \) be a source structure, let \( \varphi \in \mathcal{L} \), let \( \psi \in \mathcal{L}[\Sigma] \). The proof is by induction on the construction of \( \varphi \). For each case of the induction, we have to find a sentence \( \chi \in \mathcal{L}[\Sigma] \) such that \( \chi \equiv_{\psi} [s: \psi] \varphi \) and \( \models_{\Sigma} \chi \leftrightarrow [s: \psi] \varphi \).

Consider the case that \( \varphi \) is an atomic sentence. By the reduction law for atomic sentences, \( \models_{\Sigma} [s: \psi] \varphi \leftrightarrow \text{pre}_s(\varphi) \rightarrow \psi \), and clearly, \( \text{pre}_s(\varphi) \rightarrow \psi \equiv_{\psi} [s: \psi] \varphi \). We now assume, as our induction hypothesis, that we have shown the claim for all subformulas of \( \varphi \).

The four cases of the inductive step are all similar, so we restrict ourselves to discussing two cases: negation and defeasible knowledge.

Consider the case that \( \varphi \) is of the form \( \neg \rho \). By the reduction law for negation, \( \models_{\Sigma} [s: \psi] \neg \rho \leftrightarrow \text{pre}_s(\varphi) \rightarrow [s: \psi] \varphi \). Since \( \rho \) is a subformula of \( \varphi \), there exists a sentence \( \theta \) such that \( \models_{\Sigma} \theta \leftrightarrow [s: \psi] \varphi \) and \( \theta \equiv_{\psi} [s: \psi] \varphi \). Clearly, \( \text{pre}_s(\varphi) \rightarrow [s: \psi] \varphi \). But observe that \( \models_{\Sigma} [s: \psi] \varphi \leftrightarrow [s: \psi] \neg \rho \). Hence we have found the desired \( \chi \).

Consider now the case of defeasible knowledge. Suppose that \( \varphi \) is of the form \( \Box \rho \). By the reduction law for defeasible knowledge, \( \models_{\Sigma} [s: \psi] \Box \rho \leftrightarrow \text{pre}_s(\varphi) \rightarrow \text{pre}_s(\varphi)[s: \psi] \rho \). Since \( \rho \) is a subformula of \( \Box \rho \), by the induction hypothesis, there exists a sentence \( \beta \) such that \( \beta \equiv_{\psi} [s: \psi] \rho \). Clearly, \( \text{pre}_s(\varphi) \rightarrow [s: \psi] \rho \). But observe that \( \models_{\Sigma} [s: \psi] \rho \rightarrow \Box_{s}(\varphi, \beta) \equiv_{\psi} [s: \psi] \rho \). Hence we have found the desired \( \chi \).

Using the previous lemma, we show that any sentence in \( \mathcal{L}[\Sigma] \) can be reduced to a sentence in \( \mathcal{L} \).

**Proposition 92 ("Semantic Reduction Theorem").** For every source structure \( \Sigma \), for every sentence \( \varphi \in \mathcal{L}[\Sigma] \) there exists a sentence \( \varphi^\sharp \in \mathcal{L} \) such that \( \models_{\Sigma} \varphi \leftrightarrow \varphi^\sharp \).

**Proof.** The proof is by induction on \( c(\varphi) \). If \( c(\varphi) = 0 \), then \( \varphi \in \mathcal{L} \) (cf. Lemma 90), and we are done. Assume now that \( c(\varphi) > 0 \) and suppose that we have shown the claim for all sentences \( \varphi' \in \mathcal{L}[\Sigma] \) such that \( \varphi' \equiv_{\Sigma} \varphi \). Choose a subformula \( \psi \) of \( \varphi \) with \( h(\psi) = 1 \) (some such \( \psi \) exists, for otherwise, \( c(\varphi) = 0 \), contradiction). The sentence \( \psi \) is a Boolean combination of \( \mathcal{L} \)-sentences and sentences of the form \([s: \alpha] \theta \). Choose some such \([s: \alpha] \theta \) (again, some such
Theorem 93. For every source structure $\Sigma$: $\mathcal{L}$ and $\mathcal{L}[\Sigma]$ are co-expressive.

Proof. The fact that $\mathcal{L}[\Sigma]$ is at least as expressive as $\mathcal{L}$ follows from the fact that $\mathcal{L}[\Sigma]$ extends $\mathcal{L}$, and the semantics of $\mathcal{L}[\Sigma]$ agrees with the semantics of $\mathcal{L}$ for sentences in $\mathcal{L}$. The fact that $\mathcal{L}$ is at least as expressive as $\mathcal{L}[\Sigma]$ is a corollary of Proposition 92.

We turn to a converse of sorts to the previous theorem. Namely, we show that extending the language $\mathcal{L}$ with a non-definable dynamic attitude increases the expressive power. While this is intuitively obvious, working it out in slightly more detail serves to make the case that the notion of definability we have introduced really characterizes co-expressivity, as per the conjunction of Theorems 92 and Proposition 94 below.

Let $\tau$ be a dynamic attitude which is not definable (for short: a non-definable dynamic attitude). The language $\mathcal{L}[\tau]$ is obtained by adding a formation rule to $\mathcal{L}$ allowing us to build sentences of the form $[\tau \varphi] \psi$. The
semantics for $L[\tau]$ extends the semantics for $L$, i.e., we add the following clause:

$$J[\tau \varphi] = J[\psi]_{M^\tau[\varphi],M}.$$  

Now we observe:

**Proposition 94.** $L[\tau]$ is more expressive than $L$.

**Proof.** Since $\tau$ is not definable, there exists no sentence $\theta \in L$ such that for any plausibility model $M = (S, [],)$ and world $w \in S$: $M, w \models \theta$ iff $M^\tau[p], M, w \models \Box q$. However, $M^\tau[p], M, w \models \Box q$ iff $M, w \models (\tau p)q$. So there exists no sentence $\theta \in L$ such that for any plausibility model $M = (S, [],)$ and world $w \in S$: $M, w \models \theta$ iff $M, w \models (\tau p)q$. Hence $L[\tau]$ is more expressive than $L$.

So adding definable attitudes to $L$ yields no increase in expressive power, while adding undefinable attitudes to $L$ does yield an increase in expressive power. In this sense, the notions of definability and expressivity match.

### 5.5. Completeness

This section supplies an axiomatization of the language $L[\Sigma]$ interpreted over plausibility models.

#### 5.5.1. The Logic of Definable Sources

For every source structure $\Sigma$, the logic $L[\Sigma]$ of definable sources is given by adding the axioms below to the logic of defeasible and indefeasible knowledge $L$ (the additional axioms are just the “reduction laws” discussed above, copied from the statement of Proposition 89 for easy reference):

- $[s: \varphi] p \leftrightarrow (\text{pre}_s(\varphi) \rightarrow p),$
- $[s: \varphi] \neg \psi \leftrightarrow (\text{pre}_s(\varphi) \rightarrow \neg [s: \varphi] \psi),$
- $[s: \varphi](\varphi \land \chi) \leftrightarrow ([s: \varphi] \psi \land [s: \varphi] \chi),$
- $[s: \varphi] K \psi \leftrightarrow (\text{pre}_s(\varphi) \rightarrow K[s: \varphi] \psi),$
- $[s: \varphi] \Box \psi \leftrightarrow (\text{pre}_s(\varphi) \rightarrow \Box_s(\varphi, [s: \varphi] \psi)).$

We now show that $L[\Sigma]$ is weakly sound and complete.

**Proposition 95 (“Syntactic Reduction Theorem”).** For every source structure $\Sigma$, for every sentence $\varphi \in L[\Sigma]$: there exists a sentence $\psi \in L$ such that $\varphi$ and $\psi$ are provably equivalent in $L[\Sigma]$. 

proof. The proof is analogous to the proof of the “semantic reduction theorem” (Proposition 92): first, show the syntactic analogue of Lemma 91 using the reduction axioms. Then, prove the statement of the syntactic reduction theorem by induction on \(c(\varphi)\), using the fact that our logic allows substitution of equivalents.\(^6\) 

Theorem 96 (Completeness). For every source structure \(\Sigma\): \(L[\Sigma]\) is weakly sound and complete w.r.t. plausibility models.

proof. For soundness, it suffices to show that our axioms are valid, and that the rules of inference preserve validity. As observed earlier, the axioms of the logic of defeasible and indefeasible knowledge \(L\) are indeed valid, and the rules of inference do indeed preserve validity. Furthermore, we have shown in Proposition 89 that the reduction axioms are valid. Soundness follows by induction on the length of a derivation in \(L[\Sigma]\). For completeness, we argue as follows: suppose that \(\varphi \in L[\Sigma]\) is valid. By Proposition 95, there exists a sentence \(\varphi^l \in L\) such that \(\varphi \leftrightarrow \varphi^l\) is provable in \(L[\Sigma]\). By soundness of \(L[\Sigma]\), \(\varphi \leftrightarrow \varphi^l\) is valid, so \(\varphi^l \in L\) is valid. By completeness of \(L\), \(\varphi^l\) is provable in \(L\). Since \(L[\Sigma]\) extends \(L\), \(\varphi^l\) is also provable in \(L[\Sigma]\). Also, we know from above that \(\varphi^l \rightarrow \varphi\) is provable in \(L[\Sigma]\). So \(\varphi\) is provable in \(L\) (using modus ponens). Hence any valid sentence \(\varphi \in L[\Sigma]\) is provable in \(L[\Sigma]\). So \(L[\Sigma]\) is complete.

5.6. Logics for Mutual Trust and Distrust

In this section, we generalize our results to the multi-agent setting discussed in §2.7 and §2.8 of this dissertation. The presentation parallels the one given for the single-agent case in §§5.1–5.5 so we proceed at a faster pace.

5.6.1. Multi-Agent Plausibility Models. A (multi-agent) plausibility model is a pair 

\[ M = (\{S_a\}_{a \in A}, [\cdot]) \]

where \(\{S_a\}_{a \in A}\) is a multi-agent plausibility order, and \([\cdot]\) is a valuation.

\(^6\)Cf. Kooi (2007). The point is that if a sentence \(\varphi\) is provable in \(L[\Sigma]\), and \(\varphi\) contains an occurrence of some sentence \(\psi\) as a subformula, and, moreover, \(\psi\) and some sentence \(\psi'\) are provably equivalent in \(L[\Sigma]\), then we may replace \(\psi\) with \(\psi'\) in \(\varphi\) to obtain another sentence \(\varphi'\) that is provable in \(L[\Sigma]\).
5.6. Logics for Mutual Trust and Distrust

5.6.2. The Epistemic-Doxastic Language $\mathcal{L}[A]$. The language $\mathcal{L}[A]$ (called the (multi-agent) epistemic-doxastic language) is given by the following grammar ($p \in \Phi$):

$$
\varphi ::= p \mid \neg \varphi \mid (\varphi \land \varphi) \mid K_a \varphi \mid \Box_a \varphi,
$$

where $p \in \Phi$, and $a \in A$ ($a \neq b$). Read $K_a \varphi$ as agent $a$ infallibly (or: indefeasibly) knows that $\varphi$; read $\Box_a \varphi$ as agent $a$ defeasibly knows that $\varphi$.

5.6.3. Semantics. We interpret the language $\mathcal{L}[A]$ in the way familiar from the preceding section. Let $M = (S,[\cdot])$ be a multi-agent plausibility model. Then we define by recursion on sentence structure:

$$
\begin{align*}
[p]_M &:= [p] \cap S, \\
[-\varphi]_M &:= S \setminus [\varphi]_M, \\
[\varphi \land \psi]_M &:= [\varphi]_M \cap [\psi]_M, \\
[\Box_a \varphi]_M &:= \{ v \in S \mid v \in \Box_{S_a(o)}[\varphi]_M \}, \\
[K_a \varphi]_M &:= \{ v \in S \mid v \in K_{S_a(o)}[\varphi]_M \}.
\end{align*}
$$

As earlier, we write $M,w \vDash \varphi$ to mean that $w \in [\varphi]_M$. We say that a sentence $\varphi \in \mathcal{L}[A]$ is valid iff $[\varphi]_M = S$ for any plausibility model $M = (S,[\cdot])$. And we write $\vDash_{\mathcal{L}[A]} \varphi$ if $\varphi$ is valid.

5.6.4. The Multi-Agent Logic of Defeasible and Indefeasible Knowledge. The multi-agent logic of defeasible and indefeasible knowledge is just the obvious analogue of the single-agent logic of defeasible and indefeasible knowledge considered in the previous chapter. Formally, $L[A]$ is given by the following axioms and rules:

— Axioms:

— All instances of theorems of propositional calculus
— $K_a(\varphi \rightarrow \psi) \rightarrow (K_a \varphi \rightarrow K_a \psi)$
— $\Box_a(\varphi \rightarrow \psi) \rightarrow (\Box_a \varphi \rightarrow \Box_a \psi)$
— The S5 axioms for $K_a$
— The S4 axioms for $\Box_a$
— $K_a \varphi \rightarrow \Box_a \varphi$
— $K_a(\varphi \lor \Box_a \varphi) \land K_a(\psi \lor \Box_a \psi) \rightarrow (K_a \varphi \lor K_a \psi)$

— Rules of inference:

— From $\varphi$ and $\varphi \rightarrow \psi$ infer $\psi$
— From $\varphi$ infer $K_a \varphi$ and $\Box_a \varphi$
The axiomatization of $L[A]$ is again due to Baltag and Smets (2008).

**Theorem 97 (Baltag and Smets (2008)).** The multi-agent logic of defeasible and indefeasible knowledge $L[A]$ is weakly sound and complete w.r.t. multi-agent plausibility models.

**Proof.** Analogous to the proof of Theorem 83.

5.6.5. The Language $L[A, \Sigma]$. For each source structure $\Sigma$, we obtain the language $L[A, \Sigma]$ by adding two new construction rules to the grammar for $L[A]$, allowing us to build sentences of the form $s_{a\rightarrow b}$, with $s \in \Sigma$ a source, and $a, b \in A$, and sentences of the form $[a: \varphi]$, with $a \in A$ an agent. As detailed below, the additional syntactic material will allow us to study communication acts made by agents.

Formally, the language $L[A, \Sigma]$ is given by the following grammar:

$$\varphi ::= p \mid s_{a\rightarrow b} \mid \neg \varphi \mid (\varphi \land \varphi) \mid K_a \varphi \mid \Box_a \varphi \mid [a: \varphi]$$

where $p \in \Phi$, $a, b \in A$ ($a \neq b$), and $s \in \Sigma$.

Read $[a: \varphi]$ as “after the communication act $a: \varphi$, $\psi$ holds”, and read $s_{a\rightarrow b}$ as “agent $a$ considers agent $b$ to be a source of type $s$”. By the latter reading, we mean that the attitude of agent $a$ towards agent $b$ is given by the attitude label $\tau_s$ corresponding to the dynamic attitude $\tau_s$.

Right away, we notice that there is a “syntactic mismatch” in that a source $s := (\tau_s, \theta_s)$ over $\Sigma$ supplies us with a sentence $\tau_s \in L$, while $L[A, \Sigma]$ extends $L[A]$ (the multi-agent version) rather than $L$ (the single-agent version). This issue is easily resolved; we will take care of it in §5.6.8 below.

5.6.6. Trust-Plausibility Models over Source Structures. In what kind of structure would we like to interpret the language $L[A, \Sigma]$? The following notion is obviously useful towards answering the question.

A trust-plausibility model is a triple $((S)_{a \in A}, T, \downarrow)$ such that $((S)_{a \in A}, T)$ is a multi-agent trust-plausibility order, and $((S)_{a \in A}, \downarrow)$ is a multi-agent plausibility model.

However, we are here interested in a specific kind of trust-plausibility model. Namely, we are interested in trust-plausibility models where all the mutual dynamic attitudes of the agents come from a given source structure $\Sigma$, in the sense that if, according to the trust labeling, at some world $w$, some agent $a$ has the dynamic attitude $\sigma$ to some agent $b$, then there better exist some $s \in \Sigma$ such that $\tau_s = \sigma$. This ensures that all the attitudes agents entertain towards each other are actually definable.

So let $\Sigma$ be a source structure.
5.6. Logics for Mutual Trust and Distrust

— A trust graph \( T \) is a trust graph over \( \Sigma \) iff for each \( a, b \in A \) such that \( a \neq b \) there exists a source \( s \in \Sigma \) such that \( T(a, b) = \tau_s \).

— Let \( \{S_a\}_{a \in A} \) be a multi-agent plausibility order. A trust labeling \( T \) over \( \{S_a\}_{a \in A} \) is a trust labeling over \( \Sigma \) and \( \{S_a\}_{a \in A} \) iff for each \( w \in S: T_w \) is a trust graph over \( \Sigma \).

— A trust-plausibility model \( M = (\{S_a\}_{a \in A}, T, [\cdot]) \) is a trust-plausibility model over \( \Sigma \) iff \( T \) is a trust labeling over \( \Sigma \) and \( \{S_a\}_{a \in A} \).

A trust-plausibility model over \( \Sigma \) is thus a trust-plausibility model such that for any world \( w \) and for any pair of agents \( a, b \in A \) such that \( a \neq b \): the attitude \( T_w(a, b) \) "corresponds" to a source \( s \in \Sigma \) in the sense that \( T_w(a, b) = \tau_s \).

5.6.7. Semantics. The language \( \mathcal{L}[A, \Sigma] \) is interpreted in trust-plausibility models over \( \Sigma \).

As a preliminary: given a trust-plausibility model \( M = (\{S_a\}_{a \in A}, [\cdot], T) \), and a communication act \([a: P]\), we write \( M[a: P] \) for the trust-plausibility model \( (\{S_a[a: P]\}_{a \in A}, [\cdot], T) \).

So applying a communication act \([a: P]\) to \( M \) amounts to applying \([a: P]\) to the underlying trust-plausibility order \( \{S_a\}_{a \in A} \) and dragging the trust labeling \( T \) and the valuation \([\cdot]\) given by \( M \) along, leaving both unchanged (cf. §2.7.8 for the definition of \( S_a[a: P]\)_{a \in A}).

Now we define, by recursion on sentence structure, for each source structure \( \Sigma \), for each trust-plausibility model \( M = (\{S_a\}_{a \in A}, T, [\cdot]) \) over \( \Sigma \) and sentence \( \varphi \in \mathcal{L}[A, \Sigma] \):

\[
\begin{align*}
[p]^M & := [p] \cap S, \\
[s_{a \rightarrow b}]^M & := \{w \in S | T_w(a, b) = \tau_s\}, \\
[\neg \varphi]^M & := S \setminus [\varphi]^M, \\
[\varphi \land \psi]^M & := [\varphi]^M \cap [\psi]^M, \\
[\square_a \varphi]^M & := \{v \in S | v \in \Delta S_{a(v)}[\varphi]^M\}, \\
[K_a \varphi]^M & := \{v \in S | v \in K S_{a(v)}[\varphi]^M\}, \\
\end{align*}
\]

As before, we write \( M, w \models \varphi \) to mean that \( w \in [\varphi]^M \). We say that a sentence \( \varphi \in \mathcal{L}[A] \) is valid iff \( [\varphi]^M = S \) for any plausibility model \( M = (S, [\cdot]) \). And we write \( \models \mathcal{L}[A, \Sigma] \varphi \) if \( \varphi \) is valid.
5.6.8. Notation. For every agent \( a \), we introduce the following recursive syntactic translation \( ^a \) from sentences of \( \mathcal{L} \) to sentences of \( \mathcal{L}[A] \):

\[
p^a := p, \quad (\neg \phi)^a := \neg(\phi^a), \quad (\phi \land \psi)^a := \phi^a \land \psi^a, \quad (K \phi)^a := K_a \phi^a, \quad (\Box \phi)^a := \Box_a \phi^a.
\]

As one can see by inspecting the clauses of the definition, all that \( ^a \) does is label all occurrences of the symbols “\( K \)” and “\( \Box \)” in a given sentence with \( a \), i.e., replacing \( K \) with \( K_a \), and \( \Box \) with \( \Box_a \). We observe:

**Lemma 98.** For any \( \phi \in \mathcal{L} \), for any agent \( a \in A \), for any trust-plausibility model \( \mathcal{M} = (\{S_b\}_{b \in A}, T, []) \), for any \( w \in S \):

\[
\mathcal{M}, w \models \phi^a \iff (S_a(w), []) \models \phi
\]

**Proof.** Easy induction on \( \phi \in \mathcal{L} \). 

Next, we define abbreviations paralleling the ones we have introduced in the previous chapter (cf. §5.4.1) to the present multi-agent setting. Given a source structure \( \Sigma \), and a definition \( s = (\tau_s, \theta_s) \), recall that \( \theta_s \) is an \( \mathcal{L} \)-sentence that defines \( \tau_s \). Putting \( \theta_s^a := (\theta_s)^a \), we “personalize” \( \theta_s \) for each agent \( a \in A \).

Now we make the following abbreviations:

- \( \Diamond_b(a; \phi, \psi) := \land_{s \in \Sigma} (s_{b \rightarrow a} \rightarrow \theta^b_s(\phi, \psi)) \),
- \( \Box_b(a; \phi, \psi) := \land_{s \in \Sigma} (s_{b \rightarrow a} \rightarrow \Diamond_b(a; \phi, \psi)) \),
- \( \text{pre}(a; \phi) := \land_{s \in \Sigma} (\Diamond_b(a; \phi, \top)) \).

Again (cf. the remarks in §5.4.1), \( \Diamond_b(a; \phi, \psi) \) intentionally looks very much like a binary modality, with \( \Box_b(a; \phi, \psi) \) its “dual.” The aim is to capture, in some current trust-plausibility model, what an agent \( b \) defeasibly knows after the communication act \([a; \phi]\) is applied to that model. The sentence \( \text{pre}(a; \phi) \), on the other hand, is meant to capture the precondition (condition of executability) of the communication act \([a; \phi]\). We verify that these abbreviations fulfill their purpose in the next paragraph.

5.6.9. Reduction Laws. Establishing the desired reduction laws can now be done analogously to §5.4.3, where we considered the reduction laws for the single-agent case. Since the overall setup is somewhat different, we do provide details.

**Lemma 99.** For any source structure \( \Sigma \), for any sentence \( \phi \in \mathcal{L}[A, \Sigma] \), for any trust-plausibility model \( \mathcal{M} = (\{S_b\}_{b \in A}, T, []) \) over \( \Sigma \), for any \( w \in S \), for any \( a \in A \):

\[
[\text{pre}(a; \phi)]_{\mathcal{M}} = S[a; \phi].
\]
5.6. Logics for Mutual Trust and Distrust

**Proof.** Let $\Sigma$ be a source structure, let $\varphi \in \mathcal{L}[A, \Sigma]$, let $\mathcal{M} = (\{S_b\}_{b \in A}, T,[[\cdot]])$ be a trust-plausibility model over $\Sigma$, let $w \in S$, let $a \in A$. Consider now the following chain of equivalences:

$\mathcal{M}, w \vDash \text{pre}(a; \varphi)$

iff (unfolding the abbreviation $\text{pre}(a; \varphi)$)

$\mathcal{M}, w \vDash \bigwedge_{b \neq a} \bigwedge_{s \in \Sigma} s_{b \rightarrow a} \rightarrow \vartheta^b_s(\varphi, \top)$

iff (by the semantics and the definition of $\vartheta^b_s(\varphi, \top)$)

$\forall b \neq a \in A \forall s \in \Sigma : \text{if } T_w(b,a) = \tau_s, \text{then } \mathcal{M}[p \mapsto [[\varphi]]_M, q \mapsto W], w \vDash \vartheta^b_s(p, \tau)$

iff (by definition of $\text{pre}_s(p)$)

$\forall b \neq a \in A \forall s \in \Sigma : \text{if } T_w(b,a) = \tau_s, \text{then } (S_{b(w)}, [[\cdot]]_M, [p \mapsto [[\varphi]]_M], q \mapsto W], w \vDash \text{pre}_s(p)$

iff (by Lemma 98 and the definition of $[[\cdot]]_M$)

$\forall b \neq a \in A \forall s \in \Sigma : \text{if } T_w(b,a) = \tau_s, \text{then } w \in (S_{b(w)})^{\tau_s}_{\varphi \mapsto M}$

iff (by the fact that $\mathcal{M}$ is a trust-plausibility model over $\Sigma$)

$w \in S[a; \varphi].$

This proves our claim.

**Lemma 100.** For any source structure $\Sigma$, sentences $\varphi, \psi \in \mathcal{L}[A, \Sigma]$, for any trust-plausibility model $\mathcal{M} = (\{S_b\}_{b \in A}, T,[[\cdot]])$, world $w \in S[a; \varphi]$ and agents $a, b \in A$:

$\mathcal{M}, w \vDash [a; \varphi] \Box_b \psi \iff \mathcal{M}, w \vDash \Box_b (a; \varphi, [a; \varphi] \psi).$
Proof. Let $\Sigma$ be a source structure, let $\varphi, \psi \in \mathcal{L}(\Sigma, A)$, let $\mathcal{M} = (\{S_b\}_{b \in A}, T, [\ ])$ be a trust-plausibility model, let $w \in S[a: \varphi]$, let $a, b \in A$.

We define the trust-plausibility model $\mathcal{N}$ as

$$
\mathcal{N} := \mathcal{M}[p \mapsto [\varphi]_{\mathcal{M}}, q \mapsto [[a: \varphi]_{\mathcal{M}}].
$$

We use (cf. §2.7.7) the notation $S[a: \varphi]$ for the natural product order on $S[a: \varphi]$, that is $S[a: \varphi] := (S[a: \varphi], S[a: \varphi] \times S[a: \varphi])$. To simplify the notation, suppose that $s \in \Sigma$ is such that $T_w(b, a) = \tau_s$.

Coming from the left,

$$
\mathcal{M}, w \models [a: \varphi] \Box_b \psi
$$

iff (by the semantics and the assumption that $w \in S[a: \varphi]$)

$$
\mathcal{M}[a: \varphi], w \models \Box_b \psi
$$

iff (by the semantics and the definition of a communication act)

$$(S_{b(w)})^{\tau_s[\varphi]} \cap S[a: \varphi], w \models \Box[\varphi]_{\mathcal{M}[a: \varphi]}
$$

iff (using the definition of $S[a: \varphi]$)

$$(S_{b(w)})^{\tau_s[\varphi]}, w \models \Box(S[a: \varphi] \Rightarrow [\varphi]_{\mathcal{M}[a: \varphi]}).
$$

On the other hand, coming from the right,

$$
\mathcal{N}, w \models \Box_b (\varphi, [a: \varphi] \psi)
$$

iff (by definition of $\mathcal{N}$)

$$
\mathcal{N}, w \models \Box_b (p, q)
$$

iff (by definition of $\Box_b (p, q)$ and Lemma 98 using the fact that $T_w(b, a) = \tau_s$)

$$
\mathcal{S}_{b(w)}, w \models \Box_s (p, q)
$$

iff (since $\Box_s$ defines $\tau_s$ and $w \in S[a\varphi]$)

$$(S_{b(w)})^{\tau_s[\varphi]} \cap S[a: \varphi], w \models \Box[q]_{\mathcal{N}^s}
$$

iff (by definition of $\mathcal{N}$)

$$(S_{b(w)})^{\tau_s[\varphi]} \cap S[a: \varphi], w \models \Box[q]_{\mathcal{N}^s}.
$$

It is thus sufficient to show that

$$(S_{b(w)})^{\tau_s[\varphi]}, w \models \Box(S[a: \varphi] \Rightarrow [\varphi]_{\mathcal{M}[a: \varphi]}) \quad \text{iff} \quad (S_{b(w)})^{\tau_s[\varphi]}, w \models \Box[q]_{\mathcal{N}^s}.
$$

To prove this, we argue as follows:

$$(S_{b(w)})^{\tau_s[\varphi]}, w \models \Box(S[a: \varphi] \Rightarrow [\varphi]_{\mathcal{M}[a: \varphi]})$$
Lemma 100

We may thus assume that \( M \) and \( w \) bi-implication are satisfied at \( M \) structure, let \( \Sigma \). We consider the most interesting item, the last one. Let \( \phi \), \( \psi \), \( \chi \).

If (by the semantics)

\[
(S_{b(w)})^{\tau_s[\phi], M}, w \models \Box (S[a: \varphi] \Rightarrow (S[a: \varphi] \cap [[a: \varphi] \psi])_M)
\]

If (by propositional reasoning)

\[
(S_{b(w)})^{\tau_s[\phi], M}, w \models \Box [[a: \varphi] \psi]_M
\]

If (using Lemma 99)

\[
(S_{b(w)})^{\tau_s[\phi], M}, w \models \Box [[a: \varphi] \psi]_M
\]

If (by the semantics)

\[
(S_{b(w)})^{\tau_s[\phi], M}, w \models \Box [[a: \varphi] \psi]_M
\]

If (by definition of \( N \))

\[
(S_{b(w)})^{\tau_s[\phi], M}, w \models \Box [q]_N.
\]

This is the desired result.

\[ \Box \]

**Proposition 101.** For any source structure \( \Sigma \): the following are valid in \( L[A, \Sigma] \):

- \([a: \varphi] p \leftrightarrow (\text{pre}(a: \varphi) \rightarrow p)\)
- \([a: \varphi] s_{b \rightarrow c} \leftrightarrow (\text{pre}(a: \varphi) \rightarrow s_{b \rightarrow c})\)
- \([a: \varphi] \neg \varphi \leftrightarrow (\text{pre}(a: \varphi) \rightarrow \neg [a: \varphi] \varphi)\)
- \([a: \varphi] (\psi \land \chi) \leftrightarrow ([a: \varphi] \psi \land [a: \varphi] \chi)\)
- \([a: \varphi] K_b \psi \leftrightarrow (\text{pre}(a: \varphi) \rightarrow K_b[a: \varphi] \psi)\)
- \([a: \varphi] \Box_b \varphi \leftrightarrow (\text{pre}(a: \varphi) \rightarrow \Box_b (a: \varphi, \psi))\)

**Proof.** We consider the most interesting item, the last one. Let \( \Sigma \) be a source structure, let \( M = (\{S_a\}_{a \in A}, T, [\_]) \) be a trust-plausibility model over \( \Sigma \), and let \( w \in S \). We notice, using Lemma 99, that if \( w \notin S[a: \varphi] \), then both sides of the bi-implication are satisfied at \( w \), i.e., under this assumption, \( M, w \models [a: \varphi] \Box_b \psi \) and \( M, w \models (\text{pre}(a: \varphi) \rightarrow \Box_b (a: \varphi, [a: \varphi] \psi)) \) hold trivially, so our claim holds as well. We may thus assume that \( w \in S[a: \varphi] \), which, by Lemma 99, is equivalent to saying that \( M, w \models \text{pre}(a: \varphi) \). So we merely need to show that, under this assumption, \( M, w \models [a: \varphi] \Box_b \psi \) iff \( M, w \models \Box_b (a: \varphi, [a: \varphi] \psi) \), but this is exactly Lemma 100.
5.6.10. Logics of Mutual Trust and Distrust. For every source structure \( \Sigma \), the logic of mutual trust and distrust \( L[\Sigma, A] \) is obtained by adding the axioms below to the multi-agent logic of defeasible and indefeasible knowledge \( L[\Sigma, A] \) (the recursion axioms are just copied from the statement of Proposition[101]):

— **Recursion Axioms:**

- \([a: \varphi] p \iff (\text{pre}(a: \varphi) \to p)\)
- \([a: \varphi] s_{b\rightarrow c} \iff (\text{pre}(a: \varphi) \to s_{b\rightarrow c})\)
- \([a: \varphi] \neg \varphi \iff (\text{pre}(a: \varphi) \to \neg[a: \varphi] \varphi)\)
- \([a: \varphi](\varphi \land \chi) \iff ([a: \varphi] \varphi \land [a: \varphi] \chi)\)
- \([a: \varphi] K_b \varphi \iff (\text{pre}(a: \varphi) \to K_b[a: \varphi] \varphi)\)
- \([a: \varphi] \Box_b \varphi \iff (\text{pre}(a: \varphi) \to \Box_b(a: \varphi, [a: \varphi] \psi))\)

— **Trust Axioms:**

- \(\land_{a \in A} \land_{b \in A} (\forall s \in \Sigma s_{a\rightarrow b})\)
- \(\land_{a \in A} \land_{b \in A} \land_{s \in \Sigma} \land_{s' \in \Sigma}(s_{a\rightarrow b} \to \neg s'_{a\rightarrow b})\)

The two trust axioms capture our assumption that every agent has a unique attitude towards each other agent (cf. §2.7.5 for more discussion).

We establish the completeness of the logic of mutual trust and distrust by adapting earlier results. We provide a sketch.

**Theorem 102.** For any source structure \( \Sigma \): the logic of trust and distrust \( L[\Sigma, A] \) is weakly sound and complete w.r.t. trust-plausibility models over \( \Sigma \).

**Proof.** (Sketch.) The soundness half works as in earlier results (cf., e.g., Theorem[83]). For completeness, let \( \Sigma \) be a source structure. Our reduction axioms allow us to reduce the language \( L[A, \Sigma] \) to the language built over the grammar

\[
\varphi ::= p \mid s_{a\rightarrow b} \mid \neg \varphi \mid \varphi \land \varphi \mid K_a \varphi \mid \Box_a \varphi,
\]

with \( p \in \Phi, a \in A, s \in \Sigma \). We refer to this language as \( L[A^+] \). We refer to the proof system obtained by dropping the reduction axioms from \( L[A, \Sigma] \) as \( L[A^+] \). The fact that every sentence in \( L[A, \Sigma] \) is provably equivalent to a sentence in \( L[A^+] \) can be shown by adapting the proof of Theorem[95] using the reduction axioms. Call this adapted result the “syntactic reduction theorem for \( L[A, \Sigma] \)”.

The proof that \( L[A^+] \) is weakly sound and complete w.r.t. trust-plausibility models over \( \Sigma \) is obtained by adapting the argument in the proof of Theorem 2.5 in Baltag and Smets[2008]. Finally, we deduce, as in the proof of Theorem[66] the completeness of \( L[A^+] \) from the completeness of \( L[A, \Sigma] \), using the syntactic reduction theorem for \( L[A, \Sigma] \).