Small steps in dynamics of information

Velazquez Quesada, F.R.

Citation for published version (APA):
In the previous chapters we have explored the notions of implicit and explicit information and their dynamics, focusing in particular on the cases of implicit and explicit knowledge. But in our daily life we usually work with incomplete information, and therefore very few things are completely certain for us. The public transport in Amsterdam is highly reliable and usually behaves according to its time schedule, and nevertheless we cannot say in the absolute sense that we know the bus will be at the bus stop on time: many unpredictable factors, like flooded streets, snow, mechanical failure or even a car crash may take place. In fact, if we had to act based only on what we know, we would have very little maneuvering space. Fortunately, our attitudes toward information are more than just ‘knowing’ and ‘not knowing’. Most of our behaviour is leaded not by what we know, but rather by what we believe.

This chapter will focus on the study of the notions of implicit and explicit beliefs, as well as their dynamics. We will start by recalling two existing frameworks for representing beliefs in a possible worlds models. Then we will present our setting for representing implicit and explicit beliefs, discussing briefly some of their properties. We emphasize that, just like in the frameworks of the previous chapters there are several possibilities for defining explicit knowledge, the framework we will work with now offers us several possibilities for defining explicit beliefs (e.g., Velázquez-Quesada (2009b)). The one we will use follows the idea of defining an explicit notion of information as what is true and accepted as true in all worlds relevant for the notion (Chapters 3 and 4).

Once the static framework is settled, we will move on to discuss dynamics of implicit and explicit beliefs. First, we will recall an existing notion of belief revision in a DEL setting, refining it to put it in harmony with our non-omniscient approach. Then we will move to the new action our non-omniscient agent can perform: inferences that involve not only knowledge but also beliefs.
In order to simplify the analysis of this chapter, we will focus on the single-agent case. Moreover, we will assume that our single agent is aware of all atomic propositions of the language, and therefore aware of all formulas in the language generated by it. Nevertheless, we still keep our non-omniscient spirit: though the agent will have full attention, she does not need to recognize as true all the formulas that are so, and therefore her explicit information (in this case her explicit beliefs) does not need to be closed under logical consequence.

5.1 Approaches for representing beliefs

Let us start by reviewing two alternatives for representing beliefs within the possible worlds framework.

5.1.1 The KD45 approach

The classical approach for modelling knowledge in EL is to define it as what is true in all the worlds the agent considers possible. To get a proper representation, it is asked for the accessibility relation to be at least reflexive (making □φ → φ valid: if the agent knows φ, then φ is true), and often to be also transitive and euclidean (giving the agent full positive and negative introspection).

The idea behind the KD45 approach for representing beliefs is similar. The notion is again defined as what holds in all the worlds the agent can access from the current one, but now the accessibility relation R is asked to satisfy weaker properties. While knowledge is usually required to be true, beliefs are usually required to be just consistent. This is achieved by asking for R to be not reflexive but just serial, making the D formula ¬□⊥ valid. The additional transitivity (4) and euclideanity (5) give us full introspection.

5.1.2 Plausibility models

But beliefs are different from knowledge. Intuitively, we do not believe something because it is true in all possible situations; we believe it because it is true in the ones we consider most likely to be the case (Grove 1988; Segerberg 2001). This idea suggest that we should add further structure to the worlds that an agent consider possible. They should be given not just by a plain set, like what we get when we consider equivalence relations for representing knowledge; there should be also a plausibility order among them, indicating which worlds the agent considers more likely to be the case. This idea has led to the development of variants of possible worlds models (Board 2004; van Benthem 2007), similar to those used for conditional logics (Lewis 1973; Veltman 1985; Lamarre 1991; Boutilier 1994b). Here we recall the models we will use: the plausibility models of Baltag and Smets (2008) with small modifications.
5.1. Approaches for representing beliefs

The first question we should ask is, which properties should this plausibility order satisfy? There are several options. The minimum found in Burgess (1984) and Veltman (1985) are reflexivity and transitivity; some other authors (e.g., Lewis (1973)) impose also connectedness. There is no ideal choice, but we should be sure that the properties of our relation are enough to provide a proper definition of the notions we want to deal with.

The idea behind plausibility models is to define beliefs as what holds in the most plausible worlds. If we want consistent beliefs, we need to be sure that this set of maximal worlds is always properly defined. This can be done by asking for the relation to be a locally well-preorder.

**Definition 5.1 (Locally well-preorder)** Let \( M = \langle W, \leq \rangle \) be a possible worlds frame (that is, a possible worlds model minus the atomic valuation) in which the accessibility relation is denoted by \( \leq \).

For every world \( w \in W \), denote by \( V_w \) the comparability class of \( w \), that is, the set of worlds \( \leq \)-comparable to \( (\leq \)-above, \( \leq \)-equivalent or \( \leq \)-below) \( w \):

\[
V_w := \{ u \mid w \leq u \text{ or } u \leq w \}
\]

For every \( U \subseteq W \), denote by \( \text{Max}_{\leq}(U) \) the set of \( \leq \)-maximal worlds of \( U \), that is, the set of those worlds in \( U \) that are better than all the rest in \( U \):

\[
\text{Max}_{\leq}(U) := \{ v \in U \mid \text{for all } u \in U, u \leq v \}
\]

The accessibility relation \( \leq \) is said to be a locally well-preorder if and only if it is a preorder (a reflexive and transitive relation) such that, for each comparability class \( V_w \) and for every non-empty \( U \subseteq V_w \), the set of maximal worlds in \( U \) is non-empty, that is, \( \text{Max}_{\leq}(U) \neq \emptyset \).

Note how the existence of maximal elements in every \( U \subseteq V_w \) implies the already required reflexivity by just taking \( U \) as a singleton. Moreover, it also implies connectedness inside \( V_w \) (local connectedness): for any two-worlds set \( U = \{ w_1, w_2 \} \), the \( \text{Max}_{\leq}(U) \neq \emptyset \) requirement forces us to have \( w_1 \leq w_2 \) (so \( w_2 \) is the maximal), \( w_2 \leq w_1 \) (\( w_1 \) is the maximal), or both. In particular, if two worlds \( w_2 \) and \( w_3 \) are more plausible than a given \( w_1 \) (\( w_1 \leq w_2 \) and \( w_1 \leq w_3 \)), then these two worlds should be \( \leq \)-related (\( w_2 \leq w_3 \) or \( w_3 \leq w_2 \) or both). Finally, the requirement also implies that each comparability class is conversely well-founded, since there should be at least one element that is above the rest.

Summarizing, a locally well-preorder over a set \( W \) partitions it in one or more comparability classes, each one of them being a connected preorder that has maximal elements. In other words, a locally well-preorder is the same as a locally connected and conversely well-founded preorder. In particular, because of local connectedness, the notion of “most plausible” is global inside each comparability class, that is, the maximal worlds in each comparability class are the same from the perspective of any world belonging to it.

Now we can define what a plausibility model is.
Definition 5.2 (Plausibility model) A plausibility model is a possible worlds model $M = \langle W, \leq, V \rangle$ in which the accessibility relation, denoted now by $\leq$ and called the plausibility relation, is a locally well-preorder over $W$.

Note how, given a world $w$, the comparability class $V_w$ actually defines all the worlds the agent cannot distinguish from $w$. Of course, some worlds in $V_w$ might be less plausible than $w$ (those $u$ for which we have $u \leq w$ and $w \not\leq u$), some might be more plausible (those $u$ for which $w \leq u$ and $u \not\leq w$) and some others even equally-plausible (those satisfying both $w \leq u$ and $u \leq w$). But precisely because of that, the agent cannot rule them out if $w$ were the real one. Then, the union of $\leq$ and its converse $\geq$ gives us an equivalence relation (denoted by $\sim$) that corresponds to the agent’s epistemic indistinguishability (i.e., comparability) relation.

Before discussing the needed modalities to express beliefs our plausibility models, it is illustrative to justify the properties of the plausibility relation.

As we mentioned, reflexivity and transitivity are the minimal requirements found in the literature. Our first important choice is to allow models with multiple comparability classes instead of a single one, and the reason is that, though we will focus on the single-agent case in which a unique class is enough, considering multiple classes already will make smoother the transition to multi-agent situations, a case that is left for further analysis.

Now, why is it asked for every subset $U$ of every comparability class $V_w$ to have maximal elements instead of asking for this requirement just for every $V_w$? The reason is, again, further developments: though we will work with the notion of plain beliefs, asking for this requirement will allow us future extensions that involve the more general notion of conditional beliefs. This notion expresses not what the agent believes about the current situation, but what she would believe it was the case if she would learn that some $\psi$ was true, and plain beliefs can be defined as the particular case in which the condition $\psi$ is the always true $\top$. Now, in conditional beliefs, learning $\psi$ is understood as not considering those worlds where $\psi$ does not hold, similar to what the observation operation does (Definition 1.5). Nevertheless, while this operation allows us to express what will be the case after the learning, conditional beliefs describe what was the case before the learning took place. This is achieved by looking just at those worlds that satisfy the given $\psi$ in each comparability class, without discarding the rest of them. For this definition to work, we need to be sure that any such restriction will yield a set of worlds for which there are maximal elements; hence the requirement.

It is time to review the options we have for expressing the notion of belief. The first possibility is to work with the more general notion of conditional beliefs as a primitive by means of a modality of the form $B^{\psi} \varphi$ (“the agent believes $\varphi$ conditionally to $\psi$”), and then define the notion of belief as the particular case where the condition $\psi$ is the always true $\top$. 
5.2 Representing non-omniscient beliefs

But recall that, because of the properties of our plausibility relation, each comparability class is in fact a connected preorder that has maximal elements. Then, as observed in Stalnaker (2006) and Baltag and Smets (2008), \( \varphi \) is true in the most plausible worlds from \( w \) (i.e., the maximal ones in \( V_w \)) if and only if \( w \) can \( \leq \)-see a world from which all \( \leq \)-successors are \( \varphi \) worlds. Hence, we can use a standard modality for the relation \( \leq \) and define plain beliefs with it in the following way:

\[
B \varphi := \langle \varepsilon \rangle [\leq] \varphi
\]

As a remark, note how the notion of conditional belief cannot be defined just in terms of a modality for \( \leq \), but it can be defined if we include (1) a universal modality \( U \) and a strict plausibility modality \( \langle \prec \rangle \) or (2) a universal modality \( \langle \leq \rangle \) or (3) a modality for the epistemic indistinguishability relation \( \sim \).

5.2 Representing non-omniscient beliefs

Our framework for representing implicit and explicit beliefs combines the ideas used in previous chapters for defining implicit and explicit knowledge, with the just introduced plausibility models.

The language has two components: formulas and rules. Formulas are given by a propositional language extended, first, with formulas of the form \( A \varphi \) and \( R \rho \), where \( \varphi \) is a formula and \( \rho \) a rule (as in Chapters 2 and 4), and second, with modalities \( \langle \leq \rangle \) and \( \langle \sim \rangle \). Rules, on the other hand, are pairs consisting of a finite set of formulas, the rule’s premises, and a single formula, the rule’s conclusion (again, as in Chapters 2 and 4). The formal definition is as follows.

**Definition 5.3 (Language \( \mathcal{L} \))** Given a set of atomic propositions \( P \), formulas \( \varphi, \psi \) and rules \( \rho \) of the plausibility-access language \( \mathcal{L} \) are given, respectively, by

\[
\varphi := p \mid A \varphi \mid R \rho \mid \neg \varphi \mid \varphi \lor \psi \mid \langle \varepsilon \rangle \varphi \mid \langle \prec \rangle \varphi
\]

\[
\rho := (\{\psi_1, \ldots, \psi_n\}, \varphi)
\]

where \( p \in P \). Once again, formulas of the form \( A \varphi \) are read as “the agent has acknowledged (accepted) that formula \( \varphi \) is true”, and formulas of the form \( R \rho \) as “the agent has acknowledged (accepted) that rule \( \rho \) is truth-preserving”. For the modalities, \( \langle \varepsilon \rangle \varphi \) is read as “there is a more plausible world where \( \varphi \) holds”, and \( \langle \prec \rangle \varphi \) as “there is an epistemically indistinguishable world where \( \varphi \) holds”. Other boolean connectives as well as the universal modalities \( [\varepsilon] \) and \( [\sim] \) are defined as usual. We denote by \( \mathcal{L}_f \) the set of formulas of \( \mathcal{L} \), and by \( \mathcal{L}_r \), its set of rules.

\footnote{Note that \( [\varepsilon] \varphi \) is not adequate since it holds at \( w \) when \( \varphi \) is true in all the worlds that are more plausible than \( w \), and that includes not only the most plausible ones, but also all those laying between them and \( w \). In fact, \( [\varepsilon] \varphi \) stands for the so-called notion of safe belief.}

\footnote{\( B^V \varphi := U ((\psi \land \langle \prec \rangle \psi \rightarrow \varphi)) \); see Girard (2008).}

\footnote{\( B^V \varphi := U ((\psi \rightarrow \langle \varepsilon \rangle (\psi \land \langle \varepsilon \rangle (\varphi \land \psi \rightarrow \varphi))) \); see van Benthem and Liu (2007).}

\footnote{\( B^V \varphi := (\langle \prec \rangle \varphi \rightarrow \langle \sim \rangle (\psi \land [\varepsilon] (\psi \rightarrow \varphi))) \); see Boutilier (1994a); Baltag and Smets (2008).}
For the semantic model, we extend plausibility models with two functions, indicating the formulas and the rules the agent can access (i.e., has acknowledged as true and truth-preserving, respectively) at each possible world, just like in the semantic models of Chapters 2 and 4.

**Definition 5.4 (Plausibility-access model)** Let $P$ be a set of atomic propositions. A plausibility-access (PA) model is a tuple $M = \langle W, \leq, V, A, R \rangle$ where $(W, \leq, V)$ is a plausibility model over $P$ and

- $A : W \to \wp(L_f)$ is the **access set function**, indicating the formulas the agent has acknowledged as true (i.e., accepted) at each possible world.
- $R : W \to \wp(L_r)$ is the **rule set function**, indicating the rules the agent has acknowledged as truth-preserving (i.e., accepted) at each possible world.

Recall that if two worlds are $\leq$-related (comparable), then in fact they are epistemically indistinguishable. Then, we define the indistinguishability relation $\sim$ as the union of $\leq$ and its converse, that is, $\sim := \leq \cup \geq$. In other words, the agent cannot distinguish between two worlds if and only if she considers one of them more plausible than the other. Note that $\sim$ is different from the equal plausibility relation, which is given by the intersection between $\leq$ and $\geq$.

A pointed plausibility-access model $(M, w)$ is a plausibility-access model with a distinguished world $w \in W$.

Now for the semantic evaluation. The modalities $\langle \varepsilon \rangle$ and $\langle \sim \rangle$ are interpreted via their corresponding relation in the usual way, and formulas of the form $A \varphi$ and $R \rho$ are interpreted with our two extra functions.

**Definition 5.5 (Semantic interpretation)** Let $(M, w)$ be a pointed PA model with $M = \langle W, \leq, V, A, R \rangle$. Atomic propositions and boolean operators are interpreted as usual. For the remaining cases,

- $(M, w) \Vdash A \varphi$ iff $\varphi \in A(w)$
- $(M, w) \Vdash R \rho$ iff $\rho \in R(w)$
- $(M, w) \Vdash \langle \varepsilon \rangle \varphi$ iff there is a $u \in W$ such that $w \leq u$ and $(M, u) \Vdash \varphi$
- $(M, w) \Vdash \langle \sim \rangle \varphi$ iff there is a $u \in W$ such that $w \sim u$ and $(M, u) \Vdash \varphi$

In order to characterize syntactically the formulas in $L_f$ valid in plausibility-access models, we follow Theorem 2.5 of Baltag and Smets (2008). The important observation is that a locally well-preorder is a locally connected and conversely well-founded preorder. Then, by standard results on canonicity and modal correspondence (Chapter 4 of Blackburn et al. (2001)), the axiom system of Table 5.1 is sound and (weakly) complete for formulas of our language $\mathcal{L}$ with respect to ‘non-standard’ plausibility-access models: those in which $\leq$ is reflexive, transitive and locally connected (axioms $T_\leq$, $4_\leq$ and $LC$, respectively).
5.2. Representing non-omniscient beliefs

and \( \sim \) is the symmetric extension of \( \leq \) (axioms \( T \), \( 4 \), \( B \) and \( \text{Inc} \)). But such models have the finite model property with respect to formulas in our language, so completeness with respect to plausibility-access models follows from the fact that every finite strict preorder is conversely well-founded.

\[
\begin{array}{c|c|c}
\text{Prop} & \vdash \varphi & \text{for } \varphi \text{ a propositional tautology} \\
\hline
K_\leq & \vdash [\leq](\varphi \rightarrow \psi) \rightarrow ([\leq] \varphi \rightarrow [\leq] \psi) \\
\text{Dual}_\leq & \vdash \langle \varphi \rangle \leftrightarrow [\leq] \neg \varphi \\
\text{Nec}_\leq & \text{If } \vdash \varphi, \text{ then } \vdash [\leq] \varphi \\
\hline
T_\leq & \vdash [\leq] \varphi \rightarrow \varphi \\
4_\leq & \vdash [\leq] \varphi \rightarrow [\leq] [\leq] \varphi \\
\text{B}_\leq & \vdash \varphi \rightarrow [\leq] [\leq] \varphi \\
\hline
\text{LC} & (\langle \sim \rangle \varphi \wedge \langle \sim \rangle \psi) \rightarrow (\langle \sim \rangle (\varphi \wedge [\leq] \psi) \lor \langle \sim \rangle ([\sim] \psi \wedge [\leq] \varphi)) \\
\text{Inc} & \langle \sim \rangle \varphi \rightarrow \langle \sim \rangle \varphi \\
\end{array}
\]

Table 5.1: Axiom system for \( \mathcal{L} \) with respect to plausibility-access models.

5.2.1 Implicit and explicit beliefs, and their basic properties

It is time to define the notions of implicit and explicit beliefs. Again, just like there are several ways of defining explicit knowledge (the \( \varphi \) of Duc (1995); Jago (2006a); van Benthem (2008c) and our Chapter 2; the \( \square \varphi \wedge \varphi \) of Fagin and Halpern (1988) and van Ditmarsch and French (2009); the \( \square \varphi \) of Velázquez-Quesada (2009b); the \( \square (\varphi \wedge \varphi) \) of our Chapter 3), there are also several possibilities for defining explicit beliefs. The definitions we will use in this chapter, shown in Table 5.2 below, combine the ideas for defining beliefs as what is true in the most plausible situations (see the mentioned references in Subsection 5.1) with the ideas for defining notions of explicit information as what is true and has been recognized by the agent as true in the relevant set of worlds (see Chapters 3 and 4).

**Definition 5.6** The notions of implicit and explicit belief about formulas and rules are provided in Table 5.2. In words, the agent believes implicitly the formula \( \varphi \) (the rule \( \rho \)) if and only if \( \varphi \) (\( \text{tr} (\rho) \)) is true in the most plausible worlds, and she believes \( \varphi \) (\( \rho \)) explicitly if, in addition, she acknowledges it as true (truth-preserving) in these ‘best’ worlds.

But we also have an epistemic indistinguishability relation \( \sim \), so we can also define implicit and explicit knowledge by using its universal modality, just like we did in Chapters 3 and 4. Table 5.3 shows these definitions once again.
Chapter 5. Dynamics of implicit and explicit beliefs

The agent believes implicitly the formula $\phi$ 
\[ B_{\text{im}} \phi := \langle \leq \rangle [\leq] \phi \]

The agent believes explicitly the formula $\phi$ 
\[ B_{\text{ex}} \phi := \langle \leq \rangle [\leq] (\phi \land A \phi) \]

The agent believes implicitly the rule $\rho$ 
\[ B_{\text{im}} \rho := \langle \leq \rangle [\leq] \text{tr}(\rho) \]

The agent believes explicitly the rule $\rho$ 
\[ B_{\text{ex}} \rho := \langle \leq \rangle [\leq] (\text{tr}(\rho) \land R \rho) \]

Table 5.2: Implicit and explicit beliefs about formulas and rules.

The agent knows implicitly the formula $\phi$ 
\[ K_{\text{im}} \phi := [\sim] \phi \]

The agent knows explicitly the formula $\phi$ 
\[ K_{\text{ex}} \phi := [\sim] (\phi \land A \phi) \]

The agent knows implicitly the rule $\rho$ 
\[ K_{\text{im}} \rho := [\sim] \text{tr}(\rho) \]

The agent knows explicitly the rule $\rho$ 
\[ K_{\text{ex}} \rho := [\sim] (\text{tr}(\rho) \land R \rho) \]

Table 5.3: Implicit and explicit knowledge about formulas and rules.

The current framework allows us to represent implicit/explicit forms of knowledge/beliefs about formulas and rules. Moreover, the following validities, which follow from the contrapositive of axioms $\text{Inc}$ and $T_\leq$ of Table 5.1 ([\sim] \phi \rightarrow [\leq] \phi and $\phi \rightarrow \langle \leq \rangle \phi$, respectively), indicates that implicit and explicit knowledge imply implicit and explicit beliefs, respectively.

For formulas: $K_{\text{im}} \phi \rightarrow B_{\text{im}} \phi$ 
$K_{\text{ex}} \phi \rightarrow B_{\text{ex}} \phi$

For rules: $K_{\text{im}} \rho \rightarrow B_{\text{im}} \rho$ 
$K_{\text{ex}} \rho \rightarrow B_{\text{ex}} \rho$

Modulo the awareness notion (not considered in this chapter), the just defined notions of implicit and explicit knowledge have the properties stated in Subsection 4.3.3. Let us now review some properties of the notions of implicit/explicit belief. Again, though we will focus on the case of formulas, properties for rules can be obtained in a similar way.

The notions are global Note how the notions of implicit/explicit knowledge/beliefs are global in each comparability class. This is obvious for implicit knowledge because this notion is defined as what it is true in all the worlds the agent considers possible, that is, in all the worlds in the comparability class. Then, if the agent knows implicitly a given $\varphi$, this $\varphi$ is true in all the worlds of the comparability class, and therefore the agent knows it implicitly in any world in it. Moreover, since explicit knowledge is defined as implicit knowledge plus a requirement in all epistemically indistinguishable worlds, the notion is global too. More precisely, we have the following proposition.
### 5.2. Representing non-omniscient beliefs

**Proposition 5.1** Let \((M, w)\) be a pointed PA model. (1) If \((M, w) \vDash K_{\text{Im}} \varphi\) then, for all worlds \(u \in V_w\), \((M, u) \vDash K_{\text{Im}} \varphi\). (2) If \((M, w) \vDash K_{\text{Ex}} \varphi\) then, for all worlds \(u \in V_w\), \((M, u) \vDash K_{\text{Ex}} \varphi\). These two statements are abbreviated in the following validities:

\[
K_{\text{Im}} \varphi \rightarrow [\sim] K_{\text{Im}} \varphi \quad K_{\text{Ex}} \varphi \rightarrow [\sim] K_{\text{Ex}} \varphi
\]

But the notion of belief is also global inside each comparability class in its implicit and its explicit version. The main reason for this is that each such class is connected, and therefore even if the plausibility order branches at some point, these ramifications should converge to a topmost layer that exists because the relation is conversely well-founded.

**Proposition 5.2** Let \((M, w)\) be a pointed PA model. (1) If \((M, w) \vDash B_{\text{Im}} \varphi\) then, for all worlds \(u \in V_w\), \((M, u) \vDash B_{\text{Im}} \varphi\). (2) If \((M, w) \vDash B_{\text{Ex}} \varphi\) then, for all worlds \(u \in V_w\), \((M, u) \vDash B_{\text{Ex}} \varphi\). Again, these two statements correspond to the following validities:

\[
B_{\text{Im}} \varphi \rightarrow [\sim] B_{\text{Im}} \varphi \quad B_{\text{Ex}} \varphi \rightarrow [\sim] B_{\text{Ex}} \varphi
\]

**Basic properties** First, explicit beliefs are obviously implicit beliefs.

**Proposition 5.3** If \(\varphi\) is explicitly believed, then it is also implicitly believed, that is, the following formula is valid in PA models:

\[
B_{\text{Ex}} \varphi \rightarrow B_{\text{Im}} \varphi
\]

But, different from implicit and explicit knowledge, and though \(\leq\) is reflexive, neither implicit nor explicit beliefs have to be true. The reason is that the real world does not need to be among the most plausible ones.

**Fact 5.1** The formula \(B_{\text{Ex}} \varphi \land \sim \varphi\) is satisfiable in PA models.

Nevertheless, reflexivity makes implicit (hence explicit) beliefs consistent.

**Proposition 5.4** Implicit and explicit beliefs are consistent, that is, the following formula is valid in PA models:

\[
\neg B_{\text{Im}} \bot
\]

**Proof.** The validity can be derived with the axiom system from \(\top \rightarrow (\sim) \top\) (an instance of the contrapositive of \(T_{\leq}\), Prop, MP, \(\text{Nec}_{\leq}\) and then two applications of instances of \(\text{Dual}_{\leq}\) and MP.

**Omniscience** Implicit beliefs are omniscient.

**Proposition 5.5** All logical validities are implicitly believed and, moreover, implicit beliefs are closed under logical consequence. This gives us the following:
Chapter 5. Dynamics of implicit and explicit beliefs

- If $\varphi$ is valid, then $B_{im}\varphi$.
- $B_{im}(\varphi \rightarrow \psi) \rightarrow (B_{im}\varphi \rightarrow B_{im}\psi)$.

**Proof.** The argument for these statements is simple. For the first, if $\varphi$ is valid, then it is true in every world of every model; in particular, it is true in the most plausible worlds from any world in any model. For the second, if the most plausible worlds satisfy both $\varphi \rightarrow \psi$ and $\varphi$, then they also satisfy $\psi$.

But, again, explicit beliefs do not need to have these properties because the $A$-sets do not need to have any closure property. Nothing forces the $A$-sets to contain all validities, and having $\varphi$ and $\varphi \rightarrow \psi$ does not guarantee to have $\psi$.

**Introspection** Now let us review the introspection properties. First, implicit beliefs are positively introspective.

**Proposition 5.6** In PA-models, implicit beliefs have the positive introspection property, that is, the following formula is valid:

$$B_{im}\varphi \rightarrow B_{im}B_{im}\varphi$$

**Proof.** Here is a derivation:

- $B_{im}\varphi \rightarrow \langle \leq \rangle \leq \varphi$ by definition
- $\rightarrow \langle \leq \rangle \leq \leq \varphi$ by two applications of $4_{\leq}$
- $\rightarrow \langle \leq \rangle \leq \langle \leq \rangle \leq \varphi$ by $\leq \varphi \rightarrow \langle \leq \rangle \varphi$, derivable from $T_{\leq}$
- $\rightarrow B_{im}B_{im}\varphi$ by definition

They are also negatively introspective.

**Proposition 5.7** In PA-models, implicit beliefs have the negative introspection property, that is, the following formula is valid:

$$\neg B_{im}\varphi \rightarrow B_{im}\neg B_{im}\varphi$$

**Proof.** Here is a derivation:

- $\neg B_{im}\varphi \rightarrow \neg \langle \leq \rangle \leq \varphi$ by definition
- $\rightarrow \leq \neg \leq \varphi$ by $Dual_{\leq}$
- $\rightarrow \leq \neg \leq \leq \varphi$ by two applications of $4_{\leq}$
- $\rightarrow \langle \leq \rangle \leq \neg \langle \leq \rangle \leq \varphi$ by $\leq \varphi \rightarrow \langle \leq \rangle \varphi$
- $\rightarrow \langle \leq \rangle \neg \langle \leq \rangle \leq \varphi$ by $Dual_{\leq}$
- $\rightarrow B_{im}\neg B_{im}\varphi$ by definition
Explicit beliefs do not have these properties in the general case, again because the A-sets do not need to have any closure property. Nevertheless, we can get introspection by asking for additional requirements, like we did in Subsection 4.3.3 for the knowledge case.

For positive introspection, we need that if the agent has acknowledged that \( \varphi \) is true, then she has also acknowledged that she believes explicitly in it.

**Proposition 5.8** In PA models in which \( A \varphi \rightarrow A B_{Ex} \varphi \) is valid, explicit beliefs have the positive introspection property, that is, the following formula is valid:

\[
B_{Ex} \varphi \rightarrow B_{Ex}B_{Ex} \varphi
\]

**Proof.** Here is a derivation:

\[
\begin{align*}
B_{Ex} \varphi & \rightarrow \langle \leq \rangle [\leq] (\varphi \land A \varphi) \quad \text{by definition} \\
& \rightarrow \langle \leq \rangle [\leq] [\leq] (\varphi \land A \varphi \land A \varphi) \quad \text{by} 4_\leq \text{and Prop} \\
& \rightarrow \langle \leq \rangle [\leq] (\langle \leq \rangle [\leq] (\varphi \land A \varphi) \land [\leq] A \varphi) \quad \text{by dist. of} [\leq] \text{over} \land \\
& \rightarrow \langle \leq \rangle [\leq] [\leq] (\varphi \land A \varphi \land A \varphi) \quad \text{by} 4_\leq \text{and} T_\leq \\
& \rightarrow \langle \leq \rangle [\leq] (\langle \leq \rangle [\leq] (\varphi \land A \varphi) \land A \varphi) \quad \text{by} [\leq] \varphi \rightarrow \langle \leq \rangle \varphi \\
& \rightarrow \langle \leq \rangle [\leq] (B_{Ex} \varphi \land A \varphi) \quad \text{by definition} \\
& \rightarrow \langle \leq \rangle [\leq] (B_{Ex} \varphi \land A B_{Ex} \varphi) \quad \text{by the assumption} \\
& \rightarrow B_{Ex}B_{Ex} \varphi \quad \text{by definition} \quad \blacksquare
\]

For negative introspection, explicit belief about a given \( \varphi \) may fail because \( \varphi \) is not even implicitly believed, or because the agent has not acknowledged \( \varphi \) in the most plausible worlds. If \( \varphi \) is indeed an implicit belief, then explicit belief is negatively introspective if the agent acknowledges \( \neg B_{Ex} \varphi \) in all the best worlds every time she does not acknowledge \( \varphi \) in all of them.

**Proposition 5.9** In PA models in which \( \neg B_{Im} A \varphi \rightarrow B_{Im} A \neg B_{Ex} \varphi \) is valid, the following formula is valid:

\[
(\neg B_{Ex} \varphi \land B_{Im} \varphi) \rightarrow B_{Ex} \neg B_{Ex} \varphi
\]

**Proof.** By using the definitions, Dual_\leq and distributing [\leq] over \land, the implication’s antecedent becomes \( \langle \leq \rangle [\leq] \varphi \land \neg \langle \leq \rangle [\leq] A \varphi \), its right side being \( \neg B_{Im} A \varphi \). Then, the extra assumption gives us \( B_{Im} A \neg B_{Ex} \varphi \). But Proposition 5.10 below takes us from the left conjunct of the antecedent to \( B_{Im} \neg B_{Ex} \varphi \). Then we have \( B_{Im} \neg B_{Ex} \varphi \land B_{Im} A \neg B_{Ex} \varphi \), i.e., \( B_{Im}(\neg B_{Ex} \varphi \land A \neg B_{Ex} \varphi) \), which abbreviates as \( B_{Ex} \neg B_{Ex} \varphi \). \quad \blacksquare

Nevertheless, though in the general case explicit beliefs do not have neither positive nor negative introspection, they do have them in a weak form.
Proposition 5.10 The following formulas are valid in PA-models:

\[ B_{Ex}\varphi \rightarrow B_{Im}B_{Ex}\varphi \quad \text{and} \quad \neg B_{Ex}\varphi \rightarrow B_{Im}\neg B_{Ex}\varphi \]

Proof. Similar to the proofs of Propositions 5.6 and 5.7.

5.2.2 An example

We close the definition of the static framework for beliefs with a simple example.

Example 5.1 Consider the following plausibility-access model where the A- and the R-set of each world appears to their right in that order. The atomic propositions \( b \) and \( f \) have the reading “Chilly Willy is a bird” and “Chilly Willy flies”, respectively. In the model, (1) the agent knows explicitly that Chilly Willy is a bird and, moreover, she believes explicitly the rule stating that if it is a bird, then it flies. Nevertheless, (2) though she believes implicitly that Chilly Willy flies, (3) this belief is not explicit. All this is indicated by the formulas on the right. Since these notions are global inside each comparability class, we do not refer to some particular evaluation point.

\[ \begin{array}{c}
\begin{array}{c}
\circ \ W_2^b, f \\
(b, \{b \Rightarrow f\})
\end{array} \\
\begin{array}{c}
\circ \ W_1^b \\
(b, \{\} \}
\end{array}
\end{array} \]

(1) \( K_{Ex} b \land B_{Ex}(b \Rightarrow f) \)

(2) \( B_{Im} f \)

(3) \( \neg B_{Ex} f \)

After defining a framework for representing implicit and explicit forms of beliefs, we will now turn our attention to processes that transform them.

5.3 Belief revision

The first operation that we will review is the one that corresponds to belief revision, an action that occurs when an agent’s beliefs change in order to incorporate new external information in a consistent way (Gärdenfors 1992; Gärdenfors and Rott 1995; Williams and Rott 2001; Rott 2001). The study of this process and its properties can be traced back to the early 1980s, with the seminal work of Alchourrón et al. (1985) considered to mark the birth of the field.

Traditionally, there have been two approaches to study belief revision. The first one, what we could call the postulational approach, analyzes belief change
without committing to any fixed mechanism, proposing instead abstract general principles that a “rational” belief revision process should satisfy. Most of the initial work in the field follows this approach, with the so called AGM theory (Alchourron et al. 1985) being the most representative one. In most of these proposals, an agent’s beliefs are represented by a set of formulas closed under logical consequence (i.e., a complete theory), and in all of them three are the most relevant operations: (1) expansion of beliefs with a given \( \chi \), consisting technically in adding \( \chi \) to the set of formulas and then closing it under logical consequence; (2) contracting the beliefs with respect to \( \chi \), consisting in removing some formulas such that the closure under logical consequence of the resulting set does not contain \( \chi \); (3) revising the beliefs with respect to \( \chi \), consisting in contracting with \( \neg \chi \) and then expanding with \( \chi \).

On the other hand, some works have approached belief revision from a more constructive way, presenting concrete mechanisms that change the agent’s beliefs. Among these approaches we can mention the epistemic entrenchment functions of Gårdenfors and Makinson (1988): an ordering among formulas that indicates how strong is the agent’s belief about them, and therefore provide a way to encode factors that determine which beliefs should be discarded when revising with respect to a given \( \chi \). More interesting are the approaches that represent beliefs in a different way, like Grove (1988) which uses a structure called a system of spheres (based on the earlier work of Lewis (1973)) to construct revision functions. Like an epistemic entrenchment, a system of spheres is essentially a preorder, but now the ordered objects are no longer formulas, but complete theories.

On its most basic form, belief revision involves an agent with her beliefs, and study the way these beliefs change when new information appears. Then, it is very natural to look for a belief revision approach within the DEL framework. Here we review briefly the main idea behind the most relevant proposals.

### 5.3.1 The DEL approach

The main idea behind plausibility models is that the set of worlds the agent considers possible has in fact a further internal structure: an order indicating how plausible each possible world is. Then, an agent believes what is true in the most plausible worlds, i.e., those she considers more likely to be the case.

Now here is the key idea (van Ditmarsch 2005; van Benthem 2007; Baltag and Smets 2008): if beliefs are represented by a plausibility order, then changes in beliefs can be represented by changes in this order. In particular, the act of revising beliefs in order to accept \( \chi \) can be seen as an operation that puts \( \chi \)-worlds at the top of the plausibility order. Of course, there are several ways in which such a new order can be defined, but each one of them can be seen as a different policy for revising beliefs. Here is one of the many possibilities.
Chapter 5. Dynamics of implicit and explicit beliefs

Definition 5.7 (Upgrade operation) Let $M = \langle W, \leq, V, A, R \rangle$ be a PA model and let $\chi$ be a formula in $L_f$. The upgrade operation produces the PA model $M_{\uparrow\chi} = \langle W, \leq', V, A, R \rangle$, differing from $M$ just in the plausibility order, given by

\[
\leq' := \left( \leq; \chi ? \right) \cup \left( \neg \chi ?; \leq \right) \cup \left( \neg \chi ?; \sim; \chi ? \right)
\]

(1) \hspace{1cm} (2) \hspace{1cm} (3)

The new plausibility relation is given in a PDL style. It states that, after an upgrade with $\chi$, “all $\chi$-worlds become more plausible than all $\neg \chi$-worlds, and within the two zones, the old ordering remains” (van Benthem 2007). More precisely, in $M_{\uparrow\chi}$ we will have $w \leq' u$ if and only if in $M$ (1) $w \leq u$ and $u$ is a $\chi$-world, or (2) $w$ is a $\neg \chi$-world and $w \leq u$, or (3) $w \sim u$, $w$ is a $\neg \chi$-world and $u$ is a $\chi$-world. Again, there are many other definitions for a new plausibility relation that put $\chi$-worlds at the top (e.g., van Eijck and Wang (2008)). The presented one, so-called radical upgrade, only shows one of many options.

But not all relation-changing operations are technically adequate. We are interested in those that preserve the required model properties, and therefore keep us in the relevant class of models. In our case, we are interested in operations that do yield a locally well-preorder.

Proposition 5.11 If $M$ is a PA model, so is $M_{\uparrow\chi}$.

Proof. We need to show that if $\leq$ is a locally well-preorder, so is $\leq'$. The proof can be found in Appendix A.9.

Note the effect of the upgrade operation on the agent’s beliefs. It does not affect the $A$-sets, so we cannot expect for it to create explicit beliefs about $\chi$, since nothing guarantees that $\chi$ will be present in the most plausible worlds.

But even if we modify the definition to force $\chi$ to be present, the operation puts on top those worlds that satisfy $\chi$ in the original model $M$ (if there are none, the plausibility order will stay the same), but these worlds do not need to satisfy $\chi$ in the resulting model $M_{\uparrow\chi}$. In other words, an upgrade with $\chi$ does not necessarily make the agent believe in $\chi$, even implicitly; this is because, besides beliefs about facts, our agent also has high-order beliefs, that is, beliefs about beliefs and so on. Since the plausibility relation changes, the agent’s beliefs change, and so her beliefs about beliefs. This corresponds to the well-known Moore-like sentences (“$p$ is the case and the agent does not know it”) in Public Announcement Logic (Plaza 1989; Gerbrandy 1999) that become false after being announced, and therefore cannot be known by the agent.

Nevertheless, the operation makes the agent believe implicitly in $\chi$ if $\chi$ is a propositional formula. An upgrade does not change valuations, so if $\chi$ is purely propositional, the operation will put on top those worlds that satisfy $\chi$ in the original model $M$, and these worlds will still satisfy $\chi$ in the resulting model $M_{\uparrow\chi}$. Then, the agent will believe $\chi$ implicitly.
5.3. Belief revision

In order to represent this operation within the language, we add the existential modality $\langle \chi \uparrow \rangle$, with its universal version defined as its dual in the standard way. Formulas of the form $\langle \chi \uparrow \rangle \varphi$ are read as “it is possible for the agent to upgrade her beliefs with $\chi$ in such a way that after doing it $\varphi$ is the case”, with their semantic interpretation given in the following way.

**Definition 5.8 (Semantic interpretation)** Let $M = \langle W, \leq, V, A, R \rangle$ be a PA model and $\chi$ a formula in $L_f$. Then,

$$(M, w) \models \langle \chi \uparrow \rangle \varphi \iff (M_{\text{off}}, w) \models \varphi$$

Note how the upgrade operation is a total function: it can always be executed (there is no precondition\(^5\)) and it always yields one and only one model. Then, the semantic interpretation of the universal upgrade modality collapses to

$$(M, w) \models [\chi \uparrow] \varphi \iff (M_{\text{off}}, w) \models \varphi$$

Finally, in order to provide a sound and complete axiom system for the language with the new modality, we use reduction axioms once again.

**Theorem 5.1 (Reduction axioms for the upgrade modality)** The valid formulas of the language $L_f$ plus the upgrade modality in PA models are exactly those provable by the axioms and rules for the static base language (Table 5.1) plus the reduction axioms and modal inference rules listed in Table 5.4.

---

<table>
<thead>
<tr>
<th>Axiom</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vdash_p \langle \chi \uparrow \rangle p \leftrightarrow p$</td>
<td>Axiom for atomic propositions.</td>
</tr>
<tr>
<td>$\vdash \langle \chi \uparrow \rangle \neg \varphi \leftrightarrow \neg \langle \chi \uparrow \rangle \varphi$</td>
<td>Axiom for negation.</td>
</tr>
<tr>
<td>$\vdash \langle \chi \uparrow \rangle (\varphi \lor \psi) \leftrightarrow (\langle \chi \uparrow \rangle \varphi \lor \langle \chi \uparrow \rangle \psi)$</td>
<td>Axiom for disjunction.</td>
</tr>
<tr>
<td>$\vdash \langle \chi \uparrow \rangle (\langle \leq \rangle \varphi) \leftrightarrow (\langle \leq \rangle (\chi \land \langle \chi \uparrow \rangle \varphi) \lor (\neg \chi \land (\langle \chi \uparrow \rangle \varphi) \lor (\neg \chi \land (\langle \leq \rangle \chi \land (\langle \chi \uparrow \rangle \varphi))$</td>
<td>Axiom for indistinguishability.</td>
</tr>
<tr>
<td>$\vdash \langle \chi \uparrow \rangle (\langle \neg \rangle \varphi) \leftrightarrow \langle \neg \rangle (\langle \chi \uparrow \rangle \varphi)$</td>
<td>Axiom for negation.</td>
</tr>
<tr>
<td>$\vdash \langle \chi \uparrow \rangle$ if $\varphi$, then $\vdash [\chi \uparrow] \varphi$</td>
<td>Axiom for access.</td>
</tr>
</tbody>
</table>

---

Table 5.4: Axioms and rule for the upgrade modality.

Atomic valuation and access and rule sets are not affected by an upgrade, and the reduction axioms for atomic propositions, access and rule formulas reflect this. The reduction axioms for negation and disjunction are standard (in the case of negation recall that the operation does not have precondition), and those for the indistinguishability modality $\langle \neg \rangle$ reflects the fact that the operation

---

\(^5\)It can be argued that for the agent to upgrade her beliefs with respect to $\chi$, she needs to consider $\chi$ possible. This corresponds to $\langle \neg \rangle \chi$ as precondition for the operation.
just changes the order within each comparability class, so two worlds will be comparable after an upgrade if and only if they were comparable before.

The interesting axiom is the one for the plausibility modality $\langle \leq \rangle$. It is obtained with techniques from van Benthem and Liu (2007): if the new relation can be defined in terms of the original one with by means of a PDL-expression, $\leq' := \alpha(\leq)$, then in the new model a world $w$ can $\leq'$-reach a world that satisfies a given $\phi$, $\langle \chi \uparrow \rangle \langle \leq \rangle \phi$, if and only if in the original model $w$ could $\alpha(\leq)$-reach a world that will satisfy $\phi$ after the operation, $\langle \alpha(\leq) \rangle \langle \chi \uparrow \rangle \phi$. Then, if the PDL expression $\alpha(\leq)$ does not use iteration, the PDL axioms for sequential composition ($\langle a; b \rangle \phi \leftrightarrow \langle a \rangle \langle b \rangle \phi$), non-deterministic choice ($\langle a \cup b \rangle \phi \leftrightarrow (\langle a \rangle \phi \lor \langle b \rangle \phi$)) and test ($\langle \chi ? \rangle \phi \leftrightarrow (\chi \land \phi$)) can be successively applied to the formula $\langle \alpha(\leq) \rangle \langle \chi \uparrow \rangle \phi$ until only modalities with the original relation $\leq$ are left. In our particular case, the axiom simply translates the three-cases PDL definition of the new plausibility relation: after an upgrade with $\chi$ there is a $\leq$-reachable world where $\phi$ holds if and only if before the operation (1) there is a $\leq$-reachable $\chi$-world that will become $\phi$ after the upgrade, or (2) the current is a $\neg\chi$-world that can $\leq$-reach another that will turn into a $\phi$-one after the operation, or (3) the current is a $\neg\chi$-world that can $\sim$-reach another that is $\chi$ and will become $\phi$ after the upgrade.

A reduction axiom for the notion of conditional belief in terms of the modalities for $\leq$ and $\sim$ can be also obtained by unfolding the stated definition.

### 5.3.2 Our non-omniscient case

Recall our discussion about finer observations (Subsection 4.4.2). We emphasized that, in our approach, the agent may not have direct access to all the information each possible world provides. In other words, in general the agent has not only uncertainty about which one is the real world, but also about what holds in each one of them. So even if she considers possible a single world where some $\phi$ holds, she may not recognize it as a $\phi$-world because she may not have acknowledged that $\phi$ is indeed the case.

We already discussed how this affects the intuitive idea of an observation. Now, how does this affect the idea behind an upgrade?

The intuition behind any operation that changes the plausibility relation is that the agent rearranges the worlds according to what holds in each one of them. In particular, in the operation we defined, the agent puts the worlds she recognizes as $\chi$-worlds, that is, those where $\chi$ holds, on top of the rest of them, that is, those where $\neg\chi$ holds. But in our non-omniscient setting, the agent may not be able to tell whether a given world satisfies $\chi$ or not. Besides the worlds she identifies as $\chi$-worlds, that is, those satisfying $\chi \land A \chi$, and the worlds she identifies as $\neg\chi$-worlds, that is, those satisfying $\neg\chi \land A \neg\chi$, she can also see $\chi$-uncertain worlds, that is, those that do not satisfy neither $\chi \land A \chi$
nor $\neg \chi \land A \neg \chi$. From this perspective, the definition of the new relation is not reasonable anymore, since it assumes that the agent can identify whether $\chi$ holds or not in each possible world.

If the intuitive idea behind an upgrade operation with $\chi$ is that the agent will put on top those worlds she recognizes as $\chi$-worlds, then a non-omniscient version should reflect this. Define $G_{\chi} := \chi \land A \chi$. Then,

**Definition 5.9 (Non-omniscient upgrade operation)** Let $M = \langle W, \leq, V, A, R \rangle$ be a PA model and let $\chi$ be a formula in $L_f$. The non-omniscient upgrade operation produces the PA model $M_{\chi+\uparrow} = \langle W, \leq', V, A, R \rangle$, differing from $M$ just in the plausibility order, given now by

$$\leq' := (\leq; G_{\chi}?) \cup (\neg G_{\chi}?; \leq) \cup (\neg G_{\chi}?; \sim; G_{\chi}?)$$

In words, this revision policy states that the agent will put the worlds she recognizes as $\chi$-worlds on top of the rest of them, keeping the old ordering between the two zones. Note how, just like there are several definitions for omniscient upgrades that put $\chi$-worlds at the top, there are several definitions for non-omniscient variations that do the same with $G_{\chi}$-worlds. The one we have provided guarantees that, if $\chi$ is propositional, the agent will believe it explicitly after the upgrade, as we will discuss below.

The defined operation differs from its omniscient counterpart just in the ‘upgraded’ formula, but the structure of the new order is exactly the same. Therefore, this non-omniscient operation preserves PA models too.

Syntactically, we have the following.

**Definition 5.10 (Semantic interpretation)** Let $M = \langle W, \leq, V, A, R \rangle$ be a PA model and $\chi$ a formula in $L_f$. Then,

$$(M, w) \vDash (\chi^+ \uparrow) \varphi \iff (M_{\chi+\uparrow}, w) \vDash \varphi$$

The non-omniscient upgrade operation is a total function too, so the semantic interpretation of the universal non-omniscient upgrade modality collapses to

$$(M, w) \vDash [\chi^+ \uparrow] \varphi \iff (M_{\chi+\uparrow}, w) \vDash \varphi$$

Note now the effect of this non-omniscient upgrade operation. Again, we cannot expect for it to create even implicit beliefs about $\chi$, since once that the plausibility order has changed, the worlds that we have put on top, those that satisfied $\chi$ in the original model, may not satisfy it anymore.

---

6Now the requirement would be for the agent to consider $\chi$ explicitly possible. This corresponds to $<\sim)(\chi \land A \chi)$ as precondition for this non-omniscient operation.
But consider now the cases in which $\chi$ is a propositional formula. A non-omniscient upgrade with $\chi$ puts on top of the ordering not those worlds that satisfy $\chi$, but those worlds the agent recognizes as $\chi$ worlds, that is, worlds satisfying $\chi \land A \chi$. Then, since $\chi$ is propositional and the $A$-sets are not affected by the operation, every world satisfying $\chi \land A \chi$ in the original model will also satisfy it after the operation. Therefore, after a non-omniscient upgrade, the agent will believe $\chi$ not only implicitly, but also explicitly.

Finally, for an axiom system, the non-omniscient upgrade modality has exactly the same reduction axioms the upgrade modality has in the cases of atomic propositions, negation, disjunction, indistinguishability modality and access and rule set (Table 5.4). They key reduction axiom, the one for the plausibility relation, is simply the earlier one with $G \chi$ filled in:

$$\langle \chi \uparrow \rangle \langle \leq \rangle \phi \leftrightarrow \langle \leq \rangle \left( G \chi \land \langle \chi \uparrow \rangle \phi \right) \lor \langle \neg G \chi \land \langle \leq \rangle \langle \chi \uparrow \rangle \phi \right) \lor \langle \neg G \chi \land \langle \sim \rangle \left( G \chi \land \langle \chi \uparrow \rangle \phi \right)$$

5.4 Belief-based inference

The just discussed action, *upgrade* in its omniscient and non-omniscient versions, was borrowed from standard *DEL*. But our non-omniscient agent can perform actions omniscient agents cannot; in particular, she can perform inference. We have already a representation of this act in its *knowledge-based* form; let us review the main ideas behind it before going into the *belief-based* case.

The intuition behind the action of *knowledge-based inference* is that, if the agent knows explicitly a rule and all its premises, then an inference will make her know explicitly the rule’s conclusion. This action has been semantically defined as an operation that adds the rule’s conclusion to the $A$-set of those worlds where the agent has access to the rule and all its premises (Definition 4.9). But since the precondition of the operation is for the agent to know explicitly the rule and its premises, what the operation actually does is to add the rule’s conclusion to the $A$-set of those worlds in which the agent knows the rule and its premises.

But take a closer look at the operation. What it actually does is discard those worlds in which the agent knows explicitly the rule and all its premises, and replace them with copies that are almost identical, the only difference being that their $A$-sets now contain the conclusion of the applied rule. And this is reasonable because, under the assumption that knowledge is true information, knowledge-based inference (inference with a known rule and known premises) is simply *deductive* reasoning: the premises are true and the rule preserves the truth, so the conclusion *should* be true. In fact, knowledge-based inference can be seen as the act of recognizing two things. First, since the applied rule is truth-preserving and its premises are true, its conclusion *must* be true; second, situations where the premises are true and the rule is truth-preserving but the conclusion does not hold are not possible.
5.4. Belief-based inference

The case is different when the inference involve beliefs. Consider, for example, a situation in which the premises of a rule are explicitly known, but the rule itself is only explicitly believed. In such cases it is reasonable to consider very likely a situation in which the premises and the conclusion hold. Nevertheless, the agent should not discard a situation where the premises hold but the conclusion fails, and therefore the rule is not truth-preserving. Technically, an operation representing such action should split the current possibilities into two. One of them, the most plausible one, standing for the case in which the rule’s conclusion is indeed true; the other, the less plausible one, standing for the case in which the rule’s conclusion is false and the rule is not truth-preserving.

More generally, an inference that involves beliefs creates new possibilities, and an operation representing it should be faithful to this. But, how to do it? The action models and product update of Baltag et al. (1999), already used in Section 3.6 for multi-agent situations, will be useful once again. This time our proposal will be based in its plausibility version.

5.4.1 Plausibility-access action models

Recall the intuition behind the action models of Baltag et al. (1999): just as the agent can be uncertain about which one is the real world, she can also be uncertain about which event has taken place. In such situations, her uncertainty about the action can be represented with a model similar to that used for representing her uncertainty about the situation. Action models are possible-worlds-like structures in which the agent considers different events as possible, and her uncertainty after the action is an even combination of her uncertainty about the situation before the action and her uncertainty about the action.

This idea has been extended in order to match richer structures that indicate now only the worlds the agent considers possible but also a plausibility order among them. A first approach was made in Aucher (2003), then generalized in van Ditmarsch (2005). But these two works are based in quantitative plausibility orders that use plausibility ordinals in order to express degrees of belief. In contrast, the plausibility action models of Baltag and Smets (2008) are purely qualitative, and therefore provide a more natural extension to be used with the matching plausibility models.

Just like we did in Section 3.6, we will extend these plausibility action models in order to deal with our access and rule sets function. Here is the formal definition, for the single-agent case.

Definition 5.11 (Plausibility-access action model) A single agent plausibility-access (PA) action model is a tuple \( C = \langle E, \leq, \text{Pre}, \text{Pos}_A, \text{Pos}_R \rangle \) where

- \( \langle E, \leq, \text{Pre} \rangle \) is a plausibility action model (Baltag and Smets 2008) with \( E \) a finite non-empty set of events, \( \leq \) a plausibility order on \( E \) (with the
same requirements as those for a plausibility order in PA models) and $\text{Pre} : E \rightarrow \mathcal{L}_f$ a \textit{precondition} function indicating the requirement each event should satisfy in order to take place. This requirement is given in terms of a formula in our language $\mathcal{L}_f$, so it can include not only facts about the real world but also about the agent’s implicit/explicit knowledge/beliefs.

- $\text{Pos}_A : (E \times \wp(\mathcal{L}_f)) \rightarrow \wp(\mathcal{L}_f)$ is the \textit{new access set} function, indicating the set of formulas the agent will accept after the action according what she accepted before it and the event that has taken place.

- $\text{Pos}_R : (E \times \wp(\mathcal{L}_r)) \rightarrow \wp(\mathcal{L}_r)$ is the \textit{new rule set} function, indicating the set of rules the agent will accept after the action according what she accepted before it and the event that has taken place.

This time, we define three new relations: \textit{strict plausibility}, $\lhd := \lhd \cap \gtrsim$, \textit{equal plausibility}, $\equiv := \lhd \cap \succsim$, and \textit{epistemic indistinguishability} (i.e., comparability), $\approx := \lhd \cup \succsim$. A pointed PA action model $(C, e)$ has a distinguished event $e \in E$. ◼

Now, for the definition of the product update, note that both the static and the action model are preorders with further properties. There are two natural ways of building the order of their cartesian product: we can give the priority either the preorder of the static model, or else to that of the action model. The second option, to give priority to the order of the action, is closer to the intended spirit in which it is the \textit{action} the one that will modify the agent’s static plausibility order. The formal definition of this case is as follows.

\textbf{Definition 5.12 (Product update)} Let $M = \langle W, \leq, V, A, R \rangle$ be a PA model and $C = \langle E, \leq, \text{Pre}, \text{Pos}_A, \text{Pos}_R \rangle$ be a PA action model. The \textit{product update} operation $\otimes$ yields the PA model $M \otimes C = \langle W', \leq', V', A', R' \rangle$, given by

- $W' := \{ (w, e) \in (W \times E) \mid (M, w) \models \text{Pre}(e) \}$
- $(w_1, e_1) \leq' (w_2, e_2)$ iff $(e_1 < e_2 \text{ and } w_1 \sim w_2)$ or $(e_1 \equiv e_2 \text{ and } w_1 \leq w_2)$

and, for every $(w, e) \in W'$,

- $V'(w, e) := V(w)$
- $A'(w, e) := \text{Pos}_A(e, A(w))$
- $A'(w, e) := \text{Pos}_R(e, R(w))$ ◼

Again, the set of worlds of the new plausibility-access model is given by the restricted cartesian product of $W$ and $E$; a pair $(w, e)$ will be a world in the new model if and only if event $e$ can be executed at world $w$. Valuations and the access and rule set of the new worlds are also just as before. First, a world in the new model inherits the atomic valuation of its static component, that is,
an atom \( p \) holds at \((w, e)\) if and only if \( p \) holds at \( w \). Then, the agent’s access set at world \((w, e)\) is given by the function \( \text{Pos}_A \) with the event \( e \) and her access set at \( w \) as parameters. The case for rule sets is similar.

The important difference is that new plausibility order is now built following the so-called ‘action-priority’ rule. A world \((w_2, e_2)\) will be more plausible than \((w_1, e_1)\) if and only if either \( e_2 \) is strictly more plausible than \( e_1 \) and \( w_1 \), \( w_2 \) are already comparable (i.e., epistemically indistinguishable), or else \( e_1 \), \( e_2 \) are equally plausible and \( w_2 \) is more plausible than \( w_1 \).

Observe what our \( PA \) action models can do. If we define the new access set and new rule set functions as the identity functions (that is, \( \text{Pos}_A(e, X) := X \) and \( \text{Pos}_A(e, Y) := Y \) for all events \( e \in E \)), then we get a pure plausibility model that can modify the worlds the agent considers possible and the plausibility order among them.

But pure plausibility models cannot modify the model-component that allows us to represent finer notions of information: access and rule sets. Here is precisely where our new access set and new rule set functions, a generalization of the substitution function in van Benthem et al. (2006) for representing factual change, come into play. Our \( PA \) action models can modify the formulas and rules that the agent has acknowledged as true and truth-preserving, respectively, and therefore they can also modify the agent’s explicit beliefs.

More importantly, the main virtue of a \( PA \) action model and its product update is not that they can modify the semantic component of our static model on one hand and the syntactic component on the other. They can modify both of them together, allowing us to truly represent acts that change not only the situations the agent considers possible, but also what she has acknowledged as true in each one of them, as we will see in Subsection 5.4.2.

It is not hard to verify that product update preserves \( PA \) models.

**Proposition 5.12**. If \( M \) is a plausibility-access model and \( C \) an plausibility-access action model, then \( M \otimes C \) is a plausibility-access model.

**Proof.** We need to prove that if the plausibility orders in \( M \) and \( C \) are locally well-preorders, then so is the plausibility order of \( M \otimes C \). The proof can be found in Appendix A.10.

In order to express how product updates affect the agents’ information, we extend our language with modalities for each pointed plausibility-access action model \((C, e)\), allowing us to build formulas of the form \((C, e) \varphi\), whose semantic interpretation is given below.

**Definition 5.13 (Semantic interpretation)** Let \((M, w)\) be a pointed \( PA \) model and let \((C, e)\) be a pointed \( PA \) action model with \( \text{Pre} \) its precondition function.

\[
(M, w) \vDash (C, e) \varphi \quad \text{iff} \quad (M, w) \vDash \text{Pre}(e) \quad \text{and} \quad (M \otimes C, (w, e)) \vDash \varphi
\]
Chapter 5. Dynamics of implicit and explicit beliefs

It is now time to introduce the inferences that can be represented with PA action models.

5.4.2 Some examples of action models

Just like a public announcement in PAL corresponds to a single-event action model (Baltag et al., 1999), the action of knowledge-based inference can be represented with a single-event plausibility-access action model.

Definition 5.14 (Inference with known premises and known rule) Let $\sigma$ be a rule. The action of knowledge-based inference, that is, inference with known premises and known rule, is given by the PA action model $C_{KK}^\sigma$ whose definition (left) and diagram showing events, plausibility relation and the way rule and access sets are affected (right) are given by

\begin{align*}
E &:= \{e\} \\
\lll &= \{(e, e)\} \\
\text{Pre}(e) &:= \bigwedge_{\psi \in \text{pm}(\sigma)} K_{E_\psi} \psi \wedge K_{E_\sigma} \\
\text{Pos}_A(e, X) &:= X \cup \{\neg \text{cn}(\sigma)\} \\
\text{Pos}_R(e, Y) &:= Y \\
\end{align*}

This action model has a single event, with its precondition being for the agent to know explicitly the rule and its premises. In the resulting model, the agent will acknowledge the rule’s conclusion in all worlds satisfying the precondition. Moreover, since the premises are true and the rule is truth-preserving in all epistemically indistinguishable ($\sim$-accessible) worlds, the conclusion of the rule must be true in them in the original model $M$. But, just like the inference operation of Definition 4.9, this PA action model only affects formulas containing $\neg \text{cn}(\sigma)$; hence, $\text{cn}(\sigma)$ itself cannot be affected and will still be true in all $\sim'$-accessible worlds in the resulting model $M \otimes C_{KK}^\sigma$. Hence, the agent will know explicitly the rule’s conclusion.

But our PA action models allow us to represent more. Following our previous discussion, here is the action model for inference with known premises and believed rule.

Definition 5.15 (Inference with known premises and believed rule) Let $\sigma$ be a rule. The action of inference with known premises and believed rule is given by the PA action model $C_{KE}^\sigma$, whose definition is the following.

\begin{align*}
E &:= \{e_1, e_2\} \\
\lll &= \{(e_1, e_1), (e_1, e_2), (e_2, e_2)\} \\
\text{Pre}(e_i) &:= \bigwedge_{\psi \in \text{pm}(\sigma)} K_{E_\psi} \psi \wedge B_{E_\sigma} \\
\end{align*}

\begin{align*}
\text{Pos}_A(e_1, X) &:= X \cup \{\neg \text{cn}(\sigma)\} \\
\text{Pos}_A(e_2, X) &:= X \cup \{\text{cn}(\sigma)\} \\
\text{Pos}_R(e_1, Y) &:= Y \setminus \{\sigma\} \\
\text{Pos}_R(e_2, Y) &:= Y \\
\end{align*}
5.4. Belief-based inference

The diagram below shows this two-event model. The event on the right, the most plausible one, extends the agent’s access set with the rule’s conclusion, leaving the rule set intact; it corresponds to the case in which the conclusion of the rule holds. The one on the left, the less plausible one, extends the agent’s access set with the negation of the rule’s conclusion, removing the rule itself from the rule set; it corresponds to the case in which the conclusion of the rule does not hold. In both events the precondition is the same: the agent should know explicitly \( \sigma \)'s premises and believe explicitly \( \sigma \) itself.

We can also represent a similar situation in which the rule is explicitly known, but one or more of the premises are just explicitly believed. In this case, the best scenario is in which all believed premises are true, but one or more of them may be false, producing an extra number of situations the agent should consider. The following definition provides a model in which all these extra situations are equally plausible, but different orders can be represented.

**Definition 5.16 (Inference with believed premises and known rule)** Let \( \sigma \) be a rule, and let \( \{\psi_1, \ldots, \psi_n\} \subseteq \text{pm}(\sigma) \) be the premises of \( \sigma \) that are believed but not known. Moreover, list as \( \text{BP}_2, \ldots, \text{BP}_{2^n} \) each one of the non-empty subsets of \( \{\psi_1, \ldots, \psi_n\} \), and denote by \( \neg \text{BP}_i \) the set that contains the negation of all formulas in \( \text{BP}_i \). The action of inference with believed premises and known rule is given by the PA action model \( \mathbb{C}^\text{BK}_\sigma \) whose definition is

- \( E := \{e_1, \ldots, e_{2^n}\} \)
- \( \preceq := \{(e_i, e_1) \mid i = 1, \ldots, 2^n\} \cup \left( (E \setminus \{e_1\}) \times (E \setminus \{e_1\}) \right) \)
- \( \text{Pre}(e_i) := \bigwedge_{\psi \in \text{pm}(\sigma)} B_{\text{Ex}} \psi \land K_{\text{Ex}} \sigma \)
- \( \text{Pos}_A(e_1, X) := X \cup \{\text{cn}(\sigma)\} \)
- \( \text{Pos}_A(e_i, X) := (X \setminus \text{BP}_i) \cup \neg \text{BP}_i \) for \( i = 2, \ldots, 2^n \)
- \( \text{Pos}_R(e_i, Y) := Y \)

This time the rule is known, but some premises are just believed; then the model has one event for each combination of failing premises. Event \( e_1 \) is the one in which no premise fails, and therefore the rule’s conclusion is accepted. Events \( e_2 \) to \( e_{2^n} \) are those in which at least one premise fails, and therefore the
agent rejects them, accepting now their negation. A diagram of this \( PA \) model, with reflexive and transitive arrows omitted, appears below.

\[
\begin{align*}
E & := \{e_1, e_2\} \\
\ll & := E \times E \\
\text{Pre}(e_1) & := \text{Pre}_\sigma \\
\text{Pre}(e_2) & := \neg \text{Pre}_\sigma \\
\text{Pos}_A(e_1, X) & := X \cup \{\text{cn}(\sigma)\} \\
\text{Pos}_A(e_2, X) & := X \\
\text{Pos}_R(e_i, Y) & := Y
\end{align*}
\]
With this action the agent works *locally*. Note how any given world satisfies either the precondition of $e_1$ or the precondition of $e_2$, but not both. Then, after the operation, we will get a model that differs from the original static one only in that the agent will have accepted the rule’s conclusion exactly in those worlds in which she already accepted the rule and its premises. The diagram of this PA action model appears below.

$$X \cup \{cn(\sigma)\} \quad X$$

A stronger form of local inference can be obtained by strengthening the precondition in the following way.

**Definition 5.18 (Strong local inference)** Let $\sigma$ be a rule. The action of *strong local inference* is given by a PA action model that differs from the one representing weak local inference only in the definition of the formula $\text{Pre}_\sigma$, which is now strengthen in the following way:

$$\text{Pre}_\sigma := \left( \text{\bigwedge}_{\psi \in \text{pm}(\sigma)} (\psi \land A \psi) \right) \land (\text{tr}(\sigma) \land R \sigma)$$

The precondition of event $e_1$ now requires not only for the agent to accept the rule and the premises, but also for the premises to be true and the rule to be truth-preserving. The access sets of such worlds will be extended with the rule’s conclusion, and the rest of the worlds will remain the same.

5.4.3 A further exploration

Plausibility-access action models allow us to represent more than what we have described. We will not go into details (further applications can be found in Chapter 6), but here are some notions that arise in this rich framework.

As observed by many authors, a plausibility relation with the specified properties generates a Grove’s system of spheres [Grove 1988], that is, layers of equally-plausible elements with the layers themselves ordered according to their plausibility. The PA action models presented so far have at most two layers and in the most plausible one there are at most two events, but we do not have to restrict ourselves to them. With more than two layers we can generalize situations like the case of inference with believed premises and

---

Note how the effect of a weak local inference can be achieved with the inference operations of previous chapters (Definitions 2.16 and 4.9) by setting the appropriate precondition.
known rule: the events in which at least one of the premises fails do not need to be equally plausible. With more than two events in the top layer we can represent inferences with rules that have more than one conclusion: in the top layer we can have one event for each situation in which one or more of the conclusions are accepted.

We can also classify inferences according to how the events of the action model affect the access sets. For example, PA action models in which for every two events $e_1, e_2$ we have that $e_1 \preceq e_2$ implies $\text{Pos}_A(e_1, X) \subseteq \text{Pos}_A(e_2, X)$ reflect the optimism of the agent about the conclusion: events that extend $A$-sets are more plausible. On the opposite side we have PA action models in which for every two events $e_1, e_2$ we have that $e_1 \preceq e_2$ implies $\text{Pos}_A(e_1, X) \supseteq \text{Pos}_A(e_2, X)$; they reflect the pessimism of the agent about the conclusion since events that extend $A$-sets are less plausible.

Finally, all the inferences we have discussed follow one direction: from the rule and its premises to its conclusion. But a rule can be used in many other ways. For example, the agent can also reason by contraposition: if she knows explicitly a rule and also knows explicitly that the conclusion fails, then she can infer that at least one of the premises fails. And we do not need to stick to deductive reasoning: if the agent knows explicitly a rule and its conclusion, then she can believe explicitly that the premises hold, performing in this way a form of abductive reasoning that will be discussed in more detail in Section 6.3.

All in all, PA action models are a powerful tool that allow us to represent diverse forms of inference that involve not only an agent's knowledge but also her beliefs, therefore giving us the possibility to represent not only truth-preserving inferences but also non-truth-preserving ones (see Chapter 6). Technically, the defined product update works not only on the semantic component of the agent's information, like traditional action models do, but also on the syntactic component (formulas and rules) we have worked with through this dissertation; this allows us to truly represent acts that change not only the situations the agent considers possible, but also what she has acknowledged as true in each one of them. Thus, our agents are equipped with a broad variety of actions, and with them we can provide a more precise representation of the fine steps that changes our information in real life situations, like we will show in our example of Section 5.5.

### 5.4.4 Completeness

We have shown how PA action models can represent diverse form of inference. Let us now turn to the syntactic characterization of validities involving the PA action model modalities. Following the strategy used through this dissertation, we will provide reduction axioms for the product update operation.
The reduction axioms for atomic propositions, negation, disjunction and plausibility and indistinguishability modalities of Baltag and Smets (2008) are inherited by our system. But when looking for reduction axioms for access and rule set formulas, the functions Pos\(A\) and Pos\(R\) present a problem. The reason is that they allow the new access and rule sets to be any arbitrary set. Let us compare this with other action models and product update definitions for which reduction axioms are provided.

The action models and product update of van Benthem et al. (2006), from which our access-changing functions Pos\(A\) and Pos\(R\) and our product update have evolved, allow us to change the atomic valuation, but the new set of worlds in which each atomic proposition will be true is not arbitrary: it is given by a formula of the language. And if we see each formula as a set of worlds (those in which the formula is true), then in fact the new set of worlds in which a given atom \(p\) will be true is given in terms of the original one (the set of worlds in which \(p\) was true) by means of certain operations: ∼ (complement), ∨ (union) and so on. But not only that: the static language is already expressive enough to deal with these operations.

Let us look at another definition of an action model and its corresponding product update. The approach of van Eijck and Wang (2008) allows us to change the accessibility relation, but again the new relation is not given in an arbitrary way: it is given in terms of the previous relations by using only regular (PDL) operations. And just as the previous case, the static language is already expressive enough to deal with these expressions.

Consider now our product update operation, which extends plausibility action models by allowing us to modify sets of formulas and sets of rules. By looking at the two mentioned cases, we can see that we can provide reduction axioms in the cases in which the definitions of the Pos\(A\) and Pos\(R\) functions are not given arbitrarily, but by means of some structured expression that can be already handled in the static language. We will focus on what we will call set expressions, and here is our strategy. First, we will extend our static language in order to deal with these expressions, providing not only their semantic interpretation but also the corresponding axioms for them. Then, with the help of these new formulas, we will provide reduction axioms for the class of PA action models in which the Pos\(A\) and Pos\(R\) functions are definable by means of these expressions.

As it is currently defined, our static language allows us to look for formulas only at A- and R-sets. What we will do now is to incorporate new formulas that allow us to look not only at these basic sets, but also at more complex ones.

**Definition 5.19 (Extended \(L\))** Given a set of atomic propositions \(P\), formulas \(\varphi, \psi\), rules \(\rho\), set expressions over formulas \(\Phi, \Psi\) and set expressions over rules \(\Omega, \Upsilon\) of the extended plausibility-access language \(L\) are given, respectively, by
\( \varphi ::= p \mid \neg \varphi \mid \varphi \lor \psi \mid (\langle \rangle \varphi \mid (\langle \llangle \rangle \varphi \mid [\Phi] \varphi \mid [\Omega] \rho
\)
\( \rho ::= ([\psi_1, \ldots, \psi_n], \varphi)
\)
\( \Phi ::= A \mid \{\varphi\} \mid \Phi \cup \Psi
\)
\( \Omega ::= R \mid \{\rho\} \mid \Omega \cup \Upsilon
\)

with \( p \) an atomic proposition in \( P \).

Formulas of the form \( A \varphi \) have disappeared, leaving their place to formulas of the form \( [\Phi] \varphi \) where \( \Phi \) is what we call a set expression over formulas. While the \( A \varphi \) formulas allowed us to look only at the contents of the \( A \)-sets, formulas of the form \( [\Phi] \varphi \) allow us to look at the content of more complex sets \( \Phi \) that are built from \( A \) and singletons \( \{\varphi\} \) by means of complement and union. (The case of set expressions over rules \( \Omega \) is analogous.) We emphasize that, even though our syntax for set expressions may suggest some strong semantic content, they are just a way of making syntactic comparisons between formulas and between rules, as we will see when analyzing their axiomatization.

The behavior of the new formulas is fixed by their semantic interpretation.

**Definition 5.20 (Semantic interpretation)** Let \( (M, w) \) be a pointed \( PA \) model with \( A \) and \( R \) the access and rule sets functions, respectively. The semantic interpretation for the new formulas is given by

\[
(M, w) \vdash [A] \varphi \iff \varphi \in A(w) \quad (M, w) \vdash [R] \rho \iff \rho \in R(w)
\]
\[
(M, w) \vdash [\psi] \varphi \iff \varphi = \psi \quad (M, w) \vdash [\{\psi\}] \varphi \iff \rho \in \{\varphi\}
\]
\[
(M, w) \vdash [\Phi] \varphi \iff \varphi \not\in \Phi \quad (M, w) \vdash [\Phi \cup \Psi] \varphi \iff \varphi \in (\Phi \cup \Psi)
\]
\[
(M, w) \vdash [\Phi] \varphi \iff \varphi \not\in \Phi \quad (M, w) \vdash [\Phi \cup \Psi] \varphi \iff \varphi \in (\Phi \cup \Psi)
\]

Note, first, how \( [A] \varphi \) and \( [R] \rho \) are equivalent to the earlier \( A \varphi \) and \( R \rho \), respectively. Note also how we can even look at the contents of sets built with the intersection and difference operations following the standard definitions:

\[
\Phi \cap \Psi := \overline{\Phi} \cup \overline{\Psi} \quad \Phi \setminus \Psi := \Phi \cap \overline{\Psi}
\]

The earlier ‘static’ axiom system is not enough anymore. Though the \( A \)- and \( R \)-sets still lack any special closure property and there is still no restriction in the way they interact with each other, the additional set expressions have special behaviour, characterized by the following extra axioms.

**Theorem 5.2 (Extra axioms for extended \( \mathcal{L} \) w.r.t. \( PA \) models)** The axiom system of Table 5.1, together with the axioms from Table 5.5 is sound and (weakly) complete for the extended language \( \mathcal{L} \) with respect to plausibility-access models.

\( \blacksquare \)
5.4. Belief-based inference

\[ \text{Table 5.5: Axiom system for extended } \mathcal{L} \text{ w.r.t. plausibility-access models.} \]

\[
\begin{align*}
\text{SE}_A^{(1)} & \vdash [\{\psi\}] \psi & \text{SE}_R^{(1)} & \vdash [\{\rho\}] \rho \\
\text{SE}_A^{(1)} & \vdash -[\{\psi\}] \varphi & \text{SE}_R^{(1)} & \vdash -[\{\rho\}] \rho & \text{for } \rho \neq \varphi \\
\text{SE}_A & \vdash [\Phi] \varphi \leftrightarrow \neg[\Phi] \varphi & \text{SE}_R & \vdash [\Omega] \rho \leftrightarrow \neg[\Omega] \rho \\
\text{SE}_A^{(1)} & \vdash [\Phi \cup \Psi] \varphi \leftrightarrow ([\Phi] \varphi \lor [\Psi] \varphi) & \text{SE}_R^{(1)} & \vdash [\Omega \cup \Upsilon] \rho \leftrightarrow ([\Omega] \rho \lor [\Upsilon] \rho)
\end{align*}
\]

The new axioms reflect the behaviour of these sets operations. In the case of set expressions over formulas, axioms SE$_A^{(1)}$ indicate that $\psi$ and only $\psi$ is an element of $\{\psi\}$. Axiom SE$_A$ says that $\varphi$ is in the complement of a set if and only if it is not in the set; axiom SE$_A^{(1)}$ says that $\varphi$ is in the union of two sets if and only if it is in at least one of them. The axioms for set expressions over rules behave in a similar way.

Moreover, the axioms for complement and union actually tell us that $[\Phi] \varphi$ and $[\Phi \cup \Psi] \varphi$ are not really needed, since they can be defined as $\neg[\Phi] \varphi$ and $[\Phi] \varphi \lor [\Psi] \varphi$, respectively. In fact, from the axioms we can see that all we really need are expressions that allow us to verify syntactic identity between formulas on one side, and syntactic identity between rules on the other, like formulas of the form $[\{\psi\}] \varphi$ and $[\{\rho\}] \rho$ do (see their axioms). With such extension, our original PA language $\mathcal{L}$ (Definition 5.3) is enough for defining these new expressions. Nevertheless, we will keep this ‘syntactic sugar’ in order to make easier the reading of the formulas and, more importantly, to simplify the reduction axioms that will be provided.

With the extended language it is easy to formulate reduction axioms for PA action models that provide the new access and rule sets by means of set expressions. First, we provide a proper definition of this class.

**Definition 5.21 (SE-definable PA action model)** A set-expression (SE) definable PA action model is a PA action model in which, for each event $e$, the new access set function $\text{Pos}_A(e)$ is given by a set expression over formulas, and the new rule set function $\text{Pos}_R(e)$ is given by a set expression over rules.

Note how all the PA action models presented in Subsection 5.4.2 are SE definable. For example, in the action model for inference with known premises and believed rule (Definition 5.15), we have

- Event $e_1$: $\text{Pos}_A(e_1) := A \cup \{\neg \text{cn}(\sigma)\}$, $\text{Pos}_R(e_1) := R \setminus \{\sigma\}$.
- Event $e_2$: $\text{Pos}_A(e_2) := A \cup \{\text{cn}(\sigma)\}$, $\text{Pos}_R(e_2) := R$.

Now we can provide reduction axioms for the modalities that involve PA action models and product update.
Theorem 5.3  The axiom system built from Tables 5.1, 5.5 and Table 5.6 (with $\top$ and $\bot$ the always true and always false formula, respectively) provide a sound and (weakly) complete axiom system for formulas in the extended language $\mathcal{L}$ plus modalities for action models with respect to PA models and SE-definable PA action models.

\[
\begin{align*}
\vdash \langle C, e \rangle p & \leftrightarrow \text{Pre}(e) \land p \\
\vdash \langle C, e \rangle \neg \varphi & \leftrightarrow \text{Pre}(e) \land \neg \langle C, e \rangle \varphi \\
\vdash \langle C, e \rangle (\varphi \lor \psi) & \leftrightarrow (\langle C, e \rangle \varphi \lor \langle C, e \rangle \psi) \\
\vdash \langle C, e \rangle (\leq) \varphi & \leftrightarrow \left( \text{Pre}(e) \land \bigvee_{e \leq e'} \langle C, e' \rangle \varphi \lor \bigvee_{e \gg e'} \langle C, e'' \rangle \varphi \right) \\
\vdash \langle C, e \rangle (\sim) \varphi & \leftrightarrow \left( \text{Pre}(e) \land \bigvee_{e \sim e'} \langle C, e' \rangle \varphi \right)
\end{align*}
\]

If $\vdash \varphi$, then $\vdash [C, e] \varphi$

\[
\begin{align*}
\vdash \langle C, e \rangle [A] \varphi & \leftrightarrow \text{Pre}(e) \land \text{Pos}_A(e) \varphi \\
\vdash \langle C, e \rangle [\psi] \psi & \leftrightarrow \text{Pre}(e) \land \top \\
\vdash \langle C, e \rangle [\psi] \psi & \leftrightarrow \text{Pre}(e) \land \top \quad \text{for } \varphi \neq \psi \\
\vdash \langle C, e \rangle [\Psi] \varphi & \leftrightarrow \langle C, e \rangle \neg [\Psi] \varphi \\
\vdash \langle C, e \rangle [\Psi \cup \Phi] \varphi & \leftrightarrow \langle C, e \rangle ([\Phi] \varphi \lor [\Psi] \varphi)
\end{align*}
\]

\[
\begin{align*}
\vdash \langle C, e \rangle [R] \varphi & \leftrightarrow \text{Pre}(e) \land \text{Pos}_R(e) \varphi \\
\vdash \langle C, e \rangle [\rho] \rho & \leftrightarrow \text{Pre}(e) \land \top \\
\vdash \langle C, e \rangle [\rho] \rho & \leftrightarrow \text{Pre}(e) \land \top \quad \text{for } \rho \neq \varphi \\
\vdash \langle C, e \rangle [\Omega] \rho & \leftrightarrow \langle C, e \rangle \neg [\Omega] \rho \\
\vdash \langle C, e \rangle [\Omega \cup \gamma] \rho & \leftrightarrow \langle C, e \rangle ([\Omega] \rho \lor [\gamma] \rho)
\end{align*}
\]

Table 5.6: Axioms and rules for SE-definable action models.

On the first block, the first three axioms are standard: $\langle C, e \rangle$ does not affect atomic valuations, commute with negations (modulo the precondition) and distributes over disjunctions. The fourth, inherited from Baltag and Smets (2008), states that a $\langle C, e \rangle$ product update after which there is a more plausible $\varphi$-world can be performed if and only if the evaluation point satisfies $e$’s precondition, and in the original model there is an epistemically indistinguishable world that will satisfy $\varphi$ after a product update with a strictly more plausible $e'$, or there is a more plausible world that will satisfy $\varphi$ after a product update with an equally plausible $e''$. Finally, the fifth reduction axiom indicates that the comparability class does not change: a $\langle C, e \rangle$ product update after which there is an epistemically indistinguishable $\varphi$-world can be performed if and only if the evaluation point satisfies $e$’s precondition and there is an epistemically indistinguishable world that will satisfy $\varphi$ after a product update with an indistinguishable $e'$. 
The second block contains the axioms for set expressions over formulas, with the first one being the key. After a \((C,e)\) product update, \(\varphi\) will be in the agent’s access set if and only if \(e\)’s precondition is satisfied and \(\varphi\) is in the set expression that defines the new access set at event \(e\):

\[
\langle C, e \rangle[A] \varphi \leftrightarrow \text{Pre}(e) \land [\text{Pos}_A(e)] \varphi
\]

The simplicity of the axiom takes advantage of the fact that our extended \(\mathcal{L}\) language can deal with set expressions. As mentioned before, the original language \(\mathcal{L}\) plus expressions for syntactic identity is powerful enough to express the membership of a given formula in a set defined from A-sets and singletons by means of complement and union. Then, reduction axioms without set expressions can be provided, but we would need an inductive translation from the expression \(\text{Pos}_A(e)\) to the formula that express the membership of \(\varphi\) in it.

The remaining axioms of the second block simply unfold the static axioms for the remaining set-expressions over formulas. The third block, containing axioms for set expressions over rules, behave exactly the same.

Again, a reduction axiom for the notion of conditional belief in terms of the modalities for \(\leq\) and \(\sim\) can be obtained by unfolding the stated definition.

### 5.5 An example in motion

We close this chapter with an example of the operations we have defined.

**Example 5.2** Recall the situation of Example 5.1 whose diagram appears below on the left. The agent (1) knows explicitly that Chilly Willy is a bird and believes explicitly that if it is a bird, then it flies. Nevertheless, (2) her belief about Chilly Willy being able to fly is just implicit. The formulas below the diagram express this. We also assume that the rule \(f \Rightarrow \neg\neg f\) is present in the \(R\)-sets of both worlds, though it will not be indicated in the picture for simplicity.

Now the agent decides to use the explicitly believed rule \(b \Rightarrow f\), whose premise she knows explicitly. This corresponds to the PA action model \(C_{KB}^b\Rightarrow f\), whose diagram and precondition for both worlds appears on the right.

---

\[
\begin{align*}
\text{(1)} & \quad \text{K}_X b \land \text{B}_{X \setminus b} b \Rightarrow f \\
\text{(2)} & \quad \text{B}_{I} f \land \neg \text{B}_{E} f
\end{align*}
\]
The two worlds of the static model satisfy the precondition of the two events of the action model, so the resulting PA model has four worlds, as indicated in the diagram below.

Note how \((w_2, e_2)\) and \((w_1, e_2)\) form a copy of the original static model in which the access sets of the worlds have been extended with the rule’s conclusion, according to what event \(e_2\) indicates. The two remaining worlds, \((w_2, e_1)\) and \((w_1, e_1)\), form a copy of the original static model in which the access sets of the worlds have been extended with the negation of the rule’s conclusion and the rule has been removed from the rule sets, according to what event \(e_1\) indicates. The \(e_2\)-partition is above the \(e_1\)-partition because \(e_2\) is more plausible than \(e_1\) in the action model.

In the resulting model, (1) the agent still knows explicitly that Chilly Willy is a bird and still believes explicitly the rule stating that if it is a bird then it flies. But now (2) she also believes explicitly that it flies. Nevertheless, the rule she just applied is not known, just believed; then, conscious that the rule might fail, (3) the agent does not know neither explicitly nor implicitly that Chilly Willy flies. In fact, (4) she considers explicitly a possibility in which Chilly Willy does not fly. All this is expressed by the following formulas.

\[
\begin{align*}
(1) \quad & K_{Ex} b \land B_{Ex}(b \Rightarrow f) \\
(2) \quad & B_{Ex} f \\
(3) \quad & \neg K_{Ex} f \land \neg K_{Im} f \\
(4) \quad & \tilde{K}_{Ex} \neg f
\end{align*}
\]

where \(\tilde{K}_{Ex} \varphi\) is the ‘diamond’ version of the explicit knowledge notion \(K_{Ex} \varphi\), that is, \(\tilde{K}_{Ex} \varphi := \langle \neg \rangle (\varphi \land A \varphi)\).

While waiting for information that confirms or refutes her beliefs, our agent decides to perform weak local inference. She realizes that in the situations in which she has accepted \(f\), she should also accept \(\neg \neg f\). This action corresponds to the product update between the previous four-worlds static model (below to the left with worlds renamed) and the PA action model \(C^{f\Rightarrow\neg \neg f}\) (below to the right). The diagram of the action model includes now the different preconditions for each event, with \(Pre_{(f\Rightarrow\neg \neg f)}\) standing for the formula \(A f \land R (f \Rightarrow \neg \neg f)\).
5.5. An example in motion

The action simply extends with \( \neg \neg f \) those worlds in which the agent has accepted the rule \( f \Rightarrow \neg \neg f \) and its premise \( f \). Remember that we have assumed the rule was already present in the \( R \)-sets of the initial static model, so it is also in the \( R \)-sets of the worlds \( u_1 \) to \( u_4 \). Since \( u_3 \) and \( u_4 \) satisfy \( \text{Pre}_{(f \Rightarrow \neg \neg f)} \), they will be extended with \( \neg \neg f \), following event \( e_1 \) of the action model; since \( u_1 \) and \( u_2 \) satisfy \( \neg \text{Pre}_{(f \Rightarrow \neg \neg f)} \), they will stay the same, following event \( e_2 \). Note how, since the events are equally plausible and their preconditions are complementary, what we get is an exact copy of the static model in which the worlds that satisfy \( \text{Pre}_{(f \Rightarrow \neg \neg f)} \) are extended with \( \neg \neg f \), and worlds not satisfying it stay the same. The diagram of this resulting model appears below.

Finally, our agent gets new information: a reliable and yet fallible source tells her that in fact Chilly Willy does not fly \( (\neg f) \). Since the source is fallible, our agent should not discard those worlds where this ‘soft’ information does not hold; since the source is reliable, she should consider those satisfying the information more likely to be the case.
She can handle this information by revising her beliefs with respect to $\neg f$. The operation will put the worlds the agent recognizes as $\neg f$, that is, those satisfying $\neg f \land A \neg f$, on top of the rest, keeping the ordering in the two zones as before. In this particular case, the only world the agent recognizes as $\neg f$ is $v_1$ (formerly called $(u_1, e_2)$); then, the operation produces the following result.

\[
\begin{array}{c}
\text{(1) } K_{Ex} b \land \neg B_{Ex}(b \Rightarrow f) \land \neg B_{Im}(b \Rightarrow f) \\
\text{(2) } \neg B_{Ex} f \land \neg B_{Im} f \\
\text{(3) } \neg K_{Ex} \neg f \land \neg K_{Im} \neg f \\
\text{(4) } \neg B_{Ex} f \\
\end{array}
\]

In the resulting model, (1) the agent still knows explicitly that Chilly Willy is a bird, but now she does not believe (neither explicitly nor implicitly) anymore that if it is a bird then it flies. Moreover, (2) she does not believe (neither explicitly nor implicitly) that it flies; in fact, she believes explicitly that Chilly Willy does not fly ($\neg f$). Nevertheless, she recognizes that this does not need to be the case, and therefore (3) she does not know (neither explicitly nor implicitly) that Chilly Willy does not fly. Actually, (4) she still recognizes explicitly the possibility for it to fly.

5.6 Remarks

After previous chapters have explored some variations of the notions of implicit and explicit information with particular emphasis on the knowledge cases, this chapter has focused on the notions of implicit and explicit beliefs.

On the static side, we have provided a representation for implicit and explicit beliefs by combining the ideas for representing non-omniscient agents discussed in the previous chapters, with ideas for representing beliefs in a possible worlds setting (specifically, we have used plausibility models). Table 5.7 shows the introduced notions.
5.6. Remarks

<table>
<thead>
<tr>
<th>Notion</th>
<th>Definition</th>
<th>Model requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implicit belief about formulas.</td>
<td>$\langle \leq \rangle \varphi$</td>
<td>$\leq$ a locally well-preorder.</td>
</tr>
<tr>
<td>Implicit belief about rules.</td>
<td>$\langle \leq \rangle \text{tr}(\rho)$</td>
<td>$\leq$ a locally well-preorder</td>
</tr>
<tr>
<td>Explicit belief about formulas.</td>
<td>$\langle \leq \rangle (\varphi \land A \varphi)$</td>
<td>$\leq$ a locally well-preorder.</td>
</tr>
<tr>
<td>Explicit belief about rules.</td>
<td>$\langle \leq \rangle (\text{tr}(\rho) \land R \rho)$</td>
<td>$\leq$ a locally well-preorder.</td>
</tr>
</tbody>
</table>

Table 5.7: Static notions of information.

We have also defined, again, the notions of implicit and explicit knowledge. The definitions are the same as those of Chapter 4 minus the awareness requirement (a notion not considered in this chapter for simplicity). The main difference is that the relation that defines the notion of knowledge, the epistemic indistinguishability relation $\sim$, is not a primitive in the model anymore: it is defined as the union of the plausibility order $\leq$ and its converse $\geq$. This states that the agent cannot distinguish between two worlds if she considers one of them more plausible than the other.

On the dynamic side, we have reviewed the existing DEL approach for the act of belief revision, presenting a variant that is closer to the non-omniscient spirit of our work. But the main part of this chapter has been devoted to the study of inferences that involve beliefs. After arguing why such notion should allow the agent to create new possibilities, we have shown how the combination of existing plausibility action models with the action models that deal with the syntactic components (formulas and rules) of our extended possible worlds model (Section 3.6.2) provide us with a powerful tool that can represent different forms of inference that involve not only knowledge but also beliefs. In particular, we have introduced PA action models that represent the actions of inference with known rule and known premises, believed rule and known premises, known rule and believed premises, and strong and weak local inference. Table 5.8 summarizes the actions defined in this chapter.

Thus, our setting provides a very general perspective on the workings of inferences that mix knowledge and belief, far beyond the specifics of particular consequence relations. Though we have provided just some examples of the forms of inferences we can represent, we have by no means exhausted the possibilities. Just like the original action models allow us to represent a broad variety of announcements (public, private, hidden, etc.), our PA action models allow us to represent a broad variety of inferences, including some forms that resemble default and abductive reasoning, as we will discuss in Chapter 6.

There is an important issue in which our PA action models can shed some light. The so-called “scandal of deduction” (Hintikka 1973; Sequoiah-Grayson...
Chapter 5. Dynamics of implicit and explicit beliefs

<table>
<thead>
<tr>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upgrade</td>
<td>The agent puts on top of her order those worlds satisfying a given formula.</td>
</tr>
<tr>
<td>Non-omniscient upgrade</td>
<td>The agent puts on top of her order those worlds she recognizes as satisfying a given formula.</td>
</tr>
<tr>
<td>Knowledge-based inference</td>
<td>Inference with explicitly known premises and explicitly known rule. This is truth-preserving inference, that is, deduction.</td>
</tr>
<tr>
<td>Belief-based inference</td>
<td>Inference in which the rule or at least one of the premises is not explicitly known, but explicitly believed.</td>
</tr>
</tbody>
</table>

Table 5.8: Actions and their effects.

2008; D’Agostino and Floridi 2009) comes from the idea that deductive reasoning does not provide new information because whatever is concluded was already present in the information given by the premises. In fact, it has been argued that only non truth-preserving inferences can be considered ampliative since, if the concluded information is genuinely new, its truth cannot be guaranteed by the old information (Hintikka and Sandu 2007).

In our approach, both truth-preserving and non-truth-preserving inferences are ampliative, but in a different sense. Truth-preserving inference (i.e., deduction) is definitely ampliative because, though the agent does not get new implicit knowledge, her explicit knowledge is increased. This has already been recognized by the distinction between surface information (our explicit information) and depth information (our implicit one). More precisely, we can say that deductive inference is internally ampliative because, though it does not change the number of situations the agent considers (no change in implicit information), it does increase the information the agent has about each one of these possibilities. On the other hand, non-truth-preserving inference is ampliative in a different way: it adds more possibilities. More precisely, we can say that non-truth-preserving inference is externally ampliative because it increases the number of possibilities the agent considers.

In fact, in our syntactic-semantic setting we can see a nice interplay of four main informational activities. Hard external information, i.e., observations, remove situations the agent considered possible (the observation and announcements operations), but soft external information only rearranges them (the upgrade operations). This already happens in standard omniscient DEL, but our non-omniscient setting allows us to represent new actions, the most important one being that of inference. Our truth-preserving inference does not remove situations and does not rearrange them; what it does is to extend the informa-
5.6. Remarks

...tion the agent has about each one of them. Then, our non-truth-preserving inference is what allows the agent to generate new possible situations.

Finally, through this chapter we have interpreted the $\mathcal{A}$-sets as what the agent has acknowledged as true in each possible world, and for simplicity we have left the notion of awareness out of the picture. By incorporating the notion of awareness in this belief setting, we can provide a richer picture of the agent’s attitudes, combining what she believes with what she knows and what she is aware of. In particular, for the notion of awareness, we have now two candidates, the general notion defined in Chapter 3 or the language-based notion of Chapter 4. More importantly, in this chapter we have shown how our extended version of action models can deal with syntactic ‘acknowledgement’ dynamics, but in Chapter 3 we already showed how a similar structure can deal with syntactic ‘awareness’ dynamics. Then, just like in the static part, a combination of both can provide us with a richer setting, this time of dynamic actions that affect at the same time what the agent is paying attention to and what she acknowledges as true.