Provability Logic and Admissible Rules
Iemhoff, R.

Citation for published version (APA):

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

Tools and preliminaries

In this chapter we introduce the tools used in the following chapters of part I. In Section 3.1 we discuss some principles that are derivable in preservativity and provability logic. In Section 3.2 we discuss some basic properties of Visser’s Scheme and we prove that our formulation of the scheme is equivalent to the one used by Visser (1994). In Section 3.3 we show that all principles of the provability logic we consider are derivable in preservativity logic, and that the converse does not hold. In Section 3.4 we introduce a semantics for preservativity logic, and we define various constructions on the models given by this semantics.

3.1 Basic observations

In this section we discuss some basic principles derivable in preservativity logic. When we say that a principle is arithmetically valid we mean that all the arithmetical realizations of the principle hold. We let \( \ast \) range over arithmetical realizations. Let \( iP^- \) be the logic given by the axioms Taut, P1 P1, P2 and the rule Modus Ponens and the Preservation Rule, and let \( iK \) be the logic given by the axioms Taut, and \( K \), and the rule Modus Ponens and the Necessitation Rule (Section 3.4).

3.1.1. Lemma.

(i) for any logic \( iT \) containing \( iP^- \): \( \vdash iT A \) implies \( \vdash iT \Box A \).

(ii) \( \vdash iP- \Box (A \rightarrow B) \rightarrow A \triangleright B \) and \( \vdash iP- A \triangleright B \rightarrow (\Box A \rightarrow \Box B) \).

Proof (i) Observe that \( \vdash iT (A \rightarrow B) \) implies \( \vdash iT \top \rightarrow (A \rightarrow B) \). Hence by the Preservation Rule \( \vdash iT \top \triangleright (A \rightarrow B) \), which is equivalent to \( \Box (A \rightarrow B) \).

(ii) The second implication follows immediately from P1, using the fact that \( \Box A \) is defined as \( \top \triangleright A \). The following derivation proofs the first implication.
We have

\( \vdash_{iP^-} \Box(A \rightarrow B) \leftrightarrow \top \rightarrow (A \rightarrow B) \) \hspace{1cm} (1)

\( A \vdash \top \) \hspace{1cm} (Preservation Rule) \hspace{1cm} (2)

\( \Box(A \rightarrow B) \rightarrow A \vdash (A \rightarrow B) \) \hspace{1cm} (3)(P1) \hspace{1cm} (1)

\( A \vdash A \) \hspace{1cm} (Preservation Rule) \hspace{1cm} (4)

\( \Box(A \rightarrow B) \rightarrow A \vdash (A \wedge (A \rightarrow B)) \) \hspace{1cm} (5)(4)(P2) \hspace{1cm} (3)(4)

\( (A \wedge (A \rightarrow B)) \vdash B \) \hspace{1cm} (Preservation Rule) \hspace{1cm} (6)

\( \Box(A \rightarrow B) \rightarrow A \vdash B. \) \hspace{1cm} (5)(6)(P1)

This completes the proof. \( \Box \)

Neither \( A \vdash B \leftrightarrow \Box(A \rightarrow B) \) nor \( A \vdash B \leftrightarrow (\Box A \rightarrow \Box B) \) are arithmetically valid. For the first one, we show that if this principle would hold, then so would \( \Box \neg \neg \Box \bot \). This means that HA derives \( \neg \neg \Box \bot \). By Markov’s Rule it follows that then it derives its own inconsistency \( \Box \bot \), quod non. The following derivation shows that in the presence of \( A \vdash B \leftrightarrow \Box(A \rightarrow B) \), also \( \neg \neg \Box \bot \) is arithmetically valid.

\( \neg \Box \bot \vdash \neg \Box \bot \) \hspace{1cm} (4p)

\( \Box(\neg \Box \bot \rightarrow \neg \Box \bot) \)

\( \Box(\neg \Box \bot \rightarrow \Box \bot) \) \hspace{1cm} (L)

\( \Box(\neg \neg \Box \bot). \)

A counterexample to the second principle is given by the Rosser sentence; a consistent \( \Sigma_1 \)-sentence \( R \) such that \( \Box R \rightarrow \Box \bot \). If \( R \vdash \bot \) would hold, then by the definition of \( \vdash \), we have \( \Box(\varphi \rightarrow \bot) \) for all \( \Sigma_1 \)-sentences \( \varphi \) such that \( \Box(\varphi \rightarrow R) \) holds. Therefore, we would have \( \Box(R \rightarrow \bot) \), which contradicts the fact that \( R \) is consistent with HA.

The following lemma shows that there is an equivalent formulation of \( iP^- \) which, like \( iK \), contains the Necessitation Rule instead of the Preservation Rule. This system is the one that Visser (1994) introduced as a basic system of preservativity.

3.1.2. Lemma.

(i) The logic \( iP^- \) is equivalent to the logic consisting of the axioms \( P1, P2 \) and \( \Box(A \rightarrow B) \rightarrow A \vdash B \), and the rules Modus Ponens and Necessitation.

(ii) \( iP^- \) is conservative over \( iK \) w.r.t. formulas in the language of provability logic.
3.1. Basic observations

**Proof** (ii) follows from (i) and Lemma 3.1.1. The proof of (i) is straightforward, using the same lemma. □

The following lemma states that \( \Box \) distributes over conjunction. This is a well-known property of many modal logics.

**3.1.3. Lemma.** \( \vdash_{ip^-} (\Box A \land \Box B) \leftrightarrow \Box (A \land B) \).

**Proof** Since \( (A \land B \rightarrow A) \) and \( (A \land B \rightarrow B) \) are derivable in IPC, the Preservation Rule gives \( \vdash_{ip^-} (A \land B) \rightarrow A \) and \( \vdash_{ip^-} (A \land B) \rightarrow B \). The implication from right to left now follows by Lemma 3.1.1. For the other direction, observe \( (A \rightarrow (B \rightarrow A \land B)) \) is derivable in IPC. Thus by Lemma 3.1.1 (i) also \( \vdash_{ip^-} \Box(A \rightarrow (B \rightarrow A \land B)) \). Hence by (ii) of the same lemma, \( \vdash_{ip^-} \Box A \rightarrow \Box(B \rightarrow A \land B)) \). Applying the same step again gives \( \vdash_{ip^-} \Box A \rightarrow \Box(B \rightarrow \Box(A \land B)) \), which implies \( \vdash_{ip^-} (\Box A \land \Box B) \rightarrow \Box(A \land B) \). □

Equivalent formulas preserve the same formulas and are preserved by the same formulas, as the following lemma shows.

**3.1.4. Lemma.** For any logic \( iT \) containing \( iP^- \) we have:

if \( \vdash_{IT} A \leftrightarrow B \), then \( \vdash_{IT} C \rightarrow A \leftrightarrow C \rightarrow B \) and \( \vdash_{IT} A \rightarrow C \leftrightarrow B \rightarrow C \).

**Proof** It suffices to show that if \( \vdash_{IT} (A \rightarrow B) \), then also \( \vdash_{IT} (C \rightarrow A \rightarrow C \rightarrow B) \) and \( \vdash_{IT} (B \rightarrow C \rightarrow A \rightarrow C) \). The proof is given by the following derivation.

\[
\begin{align*}
\vdash_{IT} (A \rightarrow B) & \quad \text{by Preservation Rule implies} \\
\vdash_{IT} A \rightarrow B & \quad \text{by P1 implies} \\
\vdash_{IT} (C \rightarrow A \rightarrow C \rightarrow B) \\
\vdash_{IT} (B \rightarrow C \rightarrow A \rightarrow C). 
\end{align*}
\]

In the next lemma we state a property of preservativity logic that we will often use.

**3.1.5. Lemma.** \( \vdash_{ip^-} A \rightarrow (A \lor C) \rightarrow (B \lor C) \wedge (A \land C) \rightarrow (B \land C) \).

**Proof** Left to the reader. □

We leave it to the reader to verify that the converse of the previous lemma is also valid: the logic given by adding the principle

\[ A \rightarrow (A \lor C) \rightarrow (B \lor C) \]

to \( iP^- \) derives \( Dp \).

The next lemma contains a useful consequence of \( iP^- \).
3.1.6. Lemma. $\vdash_{iP} (A \to (B \to C)) \to (A \land B) \to C$.

Proof The proof is given by the following derivation.

$$
\begin{align*}
\vdash_{iP} & \quad (A \land B) \to B \land (A \land B) \to A & \text{(Preservation)} & (1) \\
A & \to (B \to C) \to (A \land B) \to (B \to C) & (1)(P1) & (2) \\
A & \to (B \to C) \to (A \land B) \to (B \land (B \to C)) & (1)(2)(P2) & (3) \\
(B \land (B \to C)) & \to C & \text{(Preservation)} & (4) \\
A & \to (B \to C) \to (A \land B) \to C. & (3)(4)(P1)
\end{align*}
$$

The converse of Lemma 3.1.6,

$$(A \land B) \to C \to A \to (B \to C) \quad (3.1)$$

is not arithmetically valid. A counterexample is given by $A = \top$, $B = \neg \square \bot$ and $C = \square \bot$. By 4p we have $\neg \square \bot \to \neg \square \bot$. And thus by $Lp$ and $P1$ also $\neg \square \bot \to \square \bot$. But $\square (\neg \square \bot \to \square \bot)$ does not hold, since this gives $\square \neg \square \bot$.

However, Montagna's principle shows that if in (3.1) we restrict $C$ to boxed formulas it becomes derivable in the preservativity logic of HA (and hence is arithmetically valid):

3.1.7. Lemma. $\vdash_{iPH} (A \land \square C) \to B \to A \to (\square C \to B)$.

Proof By Montagna's Principle we have that

$$
\vdash_{iPH} (A \land \square C) \to B \to (\square C \to A \land \square C) \to (\square C \to B).
$$

By the Preservation Rule it follows that

$$
\vdash_{iPH} A \to (\square C \to A \land \square C).
$$

Combining these two consequences and applying $P1$ gives,

$$
\vdash_{iPH} (A \land \square C) \to B \to A \to (\square C \to B).
$$

This completes the proof. \qed

Together with (3.1.6) the last lemma shows that (substituting $\top$ for $A$)

$$
\vdash_{iPH} \square C \to B \leftrightarrow \square (\square C \to B).
$$

Observe that 4p and $Lp$ can be replaced by equivalent principles in which only $\to$ occurs. First note that $\square A$ implies $B \to A$, for all $B$. Therefore, we can replace $\square A$
3.1. Basic observations

in 4p and Lp by B \supset A and still have arithmetically valid principles which are also derivable from our principles:

\[ \vdash_{\text{PH}} A \supset (B \supset A) \quad \vdash_{\text{PH}} ((B \supset A) \supset A) \supset A. \]

Observe that Montagna's Principle derives for all formulas \( C = \bigvee_i \bigwedge_j \Box D_{ij} \) the following formula

\[ A \supset B \rightarrow (C \supset A) \supset (C \rightarrow B). \] (3.2)

The arithmetical validity of this principle is not surprising since the arithmetical realizations of such formulas \( C \) are \( \Sigma_1 \). It is a well-known fact that HA proves completeness \( (\varphi \rightarrow \square \varphi) \), for \( \Sigma_1 \)-formulas \( \varphi \). Hence \( \Box (C \supset C) \) is in the provability logic of HA for the mentioned formulas \( C \). This follows already from (3.2): if for all \( A, B \), (3.2) is in the preservativity logic of HA then also

\[ (A \land C) \supset B \rightarrow A \supset (C \rightarrow B). \] (3.3)

Thus in particular, \( \Box (C \supset C) \) is in the provability logic: by 4p we have \( C \vdash \Box C \), and thus by (3.3) \( \vdash (C \rightarrow C) \), which is \( \Box (C \rightarrow C) \).

Noteworthy consequences

The logic \( iL \), given by \( K \) and Löb's Principle \( \Box (\Box A \supset A) \rightarrow \Box A \), derives that 'if a theory is consistent then it cannot prove that a formula is unprovable' (a slight generalization of Gödel's second incompleteness theorem):

3.1.8. Lemma. \( \vdash_{iL} \Box \neg \Box A \rightarrow \Box \perp. \)

Proof Observe that \( \neg \Box A \) implies \( (\Box A \supset A) \). By the Necessitation Rule we have \( \vdash_{iL} (\neg \Box A \supset A) \). Whence \( \vdash_{iL} \neg \Box A \supset \Box A \) by Lemma 3.1.2. In Section 4.3 we show that \( iL \) derives the principle 4. Therefore, we have \( \vdash_{iL} \Box \neg \Box A \supset \Box A \). Applying Lemma 3.1.3 gives \( \vdash_{iL} \neg \Box A \supset \Box (\neg \Box A \land \Box A) \). Hence Lemmas 3.1.3 and 3.1.2 leads to \( \vdash_{iL} \Box \neg \Box A \rightarrow \Box \perp. \) □

The logic \( iLLe \) (the logic axiomatized by \( L \) and \( Le \) over \( iK \), see Section 3.4) derives that 'if there is a proof of either \( \varphi \) or the unprovability of \( \psi \), then \( \varphi \) is provable' (note that this implies the formula in \( iL \) mentioned above):

3.1.9. Lemma. Let \( \Box \) be short for \( (B \land \Box B) \). We have

\[ (i) \quad \vdash_{iLLe} \Box (A \lor \neg \Box B) \rightarrow \Box A. \]

\[ (ii) \quad \vdash_{iLe} \Box (A \lor B) \rightarrow \Box (A \lor \Box B). \]
Proof First we prove (ii):

\begin{align*}
\vdash_{\text{iLLe}} & \Box(A \lor B) \rightarrow \Box(A \lor \Box B) \\
& \Box(A \lor B) \rightarrow \Box(A \lor B) \\
& \Box(A \lor B) \rightarrow \Box((A \lor B) \land (A \lor \Box B)) \\
& \Box(A \lor B) \rightarrow \Box(A \lor \Box B).
\end{align*}

Now the proof of (i) follows from (ii) and the fact that iL derives (\Box \neg \Box B \rightarrow \Box B), as was shown in the lemma above.

There is a consequence of the Formalized Markov Scheme that states that for formulae (\lor A_i \rightarrow \lor A_i) a stronger variant of L"ob's Principle is derivable:

3.1.10. Lemma. For \( D = (\lor A_i \rightarrow \lor A_i) \) we have

\begin{align*}
\vdash_{\text{iLMa}} & \Box(D \rightarrow \neg \neg D) \rightarrow \Box D.
\end{align*}

Proof This follows from the following derivation. Let \( D = (\lor A_i \rightarrow \lor A_i) \).

\begin{align*}
\vdash_{\text{iLMa}} & \Box(D \rightarrow \neg \neg D) \rightarrow \Box(\Box \neg \neg D \rightarrow \neg \neg D) \\
& \rightarrow \Box \neg \neg D \\
& \rightarrow \Box D.
\end{align*}

This stronger version of L is not for arbitrary \( D \) a principle of HA. For instance, HA derives \( \neg \neg(\Box \bot \lor \neg \Box \bot) \), thus also \( \Box(\neg \neg(\Box \bot \lor \neg \Box \bot)) \) by the Necessitation Rule. Therefore, HA derives \( \Box(\Box \bot \lor \neg \Box \bot) \rightarrow \neg \neg(\Box \bot \lor \neg \Box \bot) \). But it does not derive \( \Box(\Box \bot \lor \neg \Box \bot) \) as the discussion of iLLe above shows.

3.2 Remarks on Visser's Scheme

In this thesis we use a slightly different formulation of Visser's Scheme then the one used by Visser (1994). The reason for this is that when we use our formulation, the modal characterization of the scheme runs smoother. In this section we prove that the two formulations are equivalent. We also explain that outside the modal context Visser's formulation is to be preferred.

Recall that Visser's Scheme consists of the principles

\[ V_{P_n} \ (\bigwedge_{i=1}^{n} (A_i \rightarrow B_i) \rightarrow A_{n+1} \lor A_{n+2}) \vdash (\bigwedge_{i=1}^{n} A_i \rightarrow B_i)(A_1, \ldots, A_{n+2}). \]
3.2. Remarks on Visser’s Scheme

The notation \((\cdot)(\cdot)\) is given by

\[
\begin{align*}
(A)(B, C_1, \ldots, C_n) & \equiv_{df} (A)(B) \lor (A)(C_1, \ldots, C_n) \\
(A)(\bot) & \equiv_{df} \bot \\
(A)(B \land B') & \equiv_{df} (A)(B) \land (A)(B') \\
(A)(\Box B) & \equiv_{df} \Box B \\
(A)(B) & \equiv_{df} (A \to B)
\end{align*}
\]

\(B\) not of the form \(\bot, (C \land C')\) or \(\Box C\).

An equivalent formulation of Visser’s Scheme

The notation used in (Visser 1994) is the following

\[
\begin{align*}
\{A\}(B, C_1, \ldots, C_n) & \equiv_{df} \{A\}(B) \lor \{A\}(C_1, \ldots, C_n) \\
\{A\}(B) & \equiv_{df} (A)(B), \text{ for } B \text{ no disjunction or conjunction} \\
\{\cdot\}(\cdot) \text{ commutes with } \land \text{ and } \lor.
\end{align*}
\]

Note that the only difference between \((\cdot)(\cdot)\) and \(\{\cdot\}(\cdot)\) lies in the treatment of disjunctions: we have \((A)(B \lor C) = (A \to B \lor C)\) and \(\{A\}(B \lor C) = \{A\}(B) \lor \{A\}(C)\). If we replace \((\cdot)(\cdot)\) by \(\{\cdot\}(\cdot)\) in Visser’s Principles, the result is the following principle

\[
VR_n (\bigwedge_{i=1}^{n} (A_i \to B_i) \to A_{n+1}) \vdash_\Box \bigwedge_{i=1}^{n} A_i \to B_i \{A_1, \ldots, A_{n+1}\}.
\]

Let us call the scheme that consists of all the principles \(VR_n\), Visser’s Real Scheme and denote it by \(VR\). Visser (1994) has shown that Visser’s Real Scheme belongs to the preservativity logic of HA. In the next proposition we show that Visser’s Scheme and Visser’s Real Scheme are interderivable. We need \((ii)\) of the following lemma. Part \((i)\) of the lemma will be used in other chapters.

3.2.1. Lemma.

\((i)\) \((A)(B)\) implies \((A \to B)\), and \((A)(B) \lor (A)(C)\) implies \((A)(B \lor C)\).

\((ii)\) For \(A = (A_1 \wedge \ldots \wedge A_{n+1})\), for all \(m\), we have

\[
\vdash_\Box PV (A \to A_{n+1} \lor \ldots \lor A_{n+m}) \vdash_\Box (A)(A_1, \ldots, A_{n+m}).
\]
Proof (i) Left to the reader; for the first statement, use induction on \( B \), for the second statement, use the first one.

(ii) Use induction on \( m \). For \( m = 1 \), observe that \( (A \rightarrow A_{n+1}) \) is equivalent to \( (A \rightarrow A_{n+1} \lor \bot) \). We leave the rest of this case to the reader. For \( m = 2 \) the statement holds by the definition of Visser's Scheme. For \( m > 2 \), we let \( C = A_{n+2} \lor \ldots \lor A_{n+m} \). It is clear that

\[
\vdash \Pi^V (A \rightarrow A_{n+1} \lor \ldots \lor A_{n+m}) \vdash (A \rightarrow A_{n+1} \lor C).
\]

By the definition of Visser's Scheme we have that

\[
\vdash \Pi^V (A \rightarrow A_{n+1} \lor C) \vdash (A)(A_1, \ldots, A_{n+1}, C).
\]

Note that because \( C \) is a disjunction it holds that \((A)(C) = (A \rightarrow C)\). By induction hypothesis we have

\[
\vdash \Pi^V (A \rightarrow C) \vdash (A)(A_1, \ldots, A_n, A_{n+2}, \ldots, A_{n+m}).
\]

We leave it to the reader to check that, using the Disjunctive Principle and \( P1 \), all this leads to the desired result,

\[
\vdash \Pi^V (A \rightarrow A_{n+1} \lor \ldots \lor A_{n+m}) \vdash (A)(A_1, \ldots, A_{n+1}).
\]

\( \Box \)

3.2.2. Proposition. Visser's Scheme derives Visser's Real Scheme and vice versa.

Proof We leave the proof that Visser's Real Scheme derives Visser's Scheme to the reader (use the fact that \( \{ A \}(B) \) implies \( (A)(B) \)). For the other part, consider a formula \((A \rightarrow A_{n+1})\), where \( A = \bigwedge_{i=1}^{n} (A_i \rightarrow B_i) \). We have to show that

\[
\vdash \Pi^V (A \rightarrow A_{n+1}) \vdash \{ A \}(A_1, \ldots, A_{n+1}). \tag{3.4}
\]

It is easy to see that every \( A_i \) is equivalent to a formula of the form

\[
A'_i = \bigvee_{j=1}^{k_i} \bigwedge_{h=1}^{m_{ij}} (C_{ijh} \land \Box D_{ijh}),
\]

where every \( C_{ijh} \) is a propositional variable, an implication or a preservation that is not a boxed formula, and such that for every \( E \), \( \{ E \}(A_i) \) is equivalent to \( \{ E \}(A'_i) \). Namely, \( A'_i \) can be obtained by replacing, in \( A_i \), occurrences \((B \lor C) \land D\) by \((B \land D) \lor (C \land D)\).

Observe that \( A \) is equivalent to \( A' \), where

\[
A' = \bigwedge_{i=1}^{n} \bigwedge_{j=1}^{k_i} (C_{ij} \land \Box D_{ij} \rightarrow B_i)).
\]
By Lemma 3.2.1 and the definition of $(\cdot)(\cdot)$ we have
\[ \vdash_{iPV} (A' \rightarrow A'_{n+1}) \supseteq \bigvee_{i=1}^{n+1} \bigwedge_{j} (\bigwedge_{h} ((A' \rightarrow C_{i,j,h}) \land D_{i,j,h})). \]
Since clearly, \( \{A';(A'_i) = \bigwedge_{i} (\bigwedge_{j} ((A' \rightarrow C_{i,j,h}) \land D_{i,j,h})) \), this implies that
\[ \vdash_{iPV} (A' \rightarrow A'_{n+1}) \supseteq \{A'(A'_1, \ldots, A'_{n+1}) \). \]
As we just observed, \( \{A'(A') \) is equivalent to \( \{A'(A_i) \). Moreover, A is equivalent to \( A' \). Thus we can conclude (3.4), and we are done.

The logic given by Visser’s Scheme

In contrast to the other principles of preservativity logic, Visser’s Scheme consists of infinitely many principles; in Section 4.6.1 we will prove that it cannot be reduced to one principle. However, also in another respect Visser’s Scheme deviates from the other principles of iPH. Namely, for all of these principles it is trivial to see that the set of all (substitution) instances of the principle is closed under substitution, and hence the logic given by the principle is closed under substitution. For Visser’s Scheme the latter holds but the former does not. Consider for example the following two instances of Visser’s Scheme:

\[
((p_1 \rightarrow q) \rightarrow p_2 \lor p_3) \supset ((p_1 \rightarrow q) \rightarrow p_1) \lor ((p_1 \rightarrow q) \rightarrow p_2) \lor ((p_1 \rightarrow q) \rightarrow p_3). \tag{3.5}
\]

\[
((\Box p_1 \rightarrow q) \rightarrow \Box p_2 \lor \Box p_3) \supset (\Box p_1 \lor \Box p_2 \lor \Box p_3). \tag{3.6}
\]

If we substitute \( \Box p_i \) for \( p_i \) in (3.5) we arrive at the formula

\[
((\Box p_1 \rightarrow q) \rightarrow \Box p_2 \lor \Box p_3) \supset ((\Box p_1 \rightarrow q) \rightarrow \Box p_1) \lor ((\Box p_1 \rightarrow q) \rightarrow \Box p_2) \lor ((\Box p_1 \rightarrow q) \rightarrow \Box p_3). \tag{3.7}
\]

This formula is not an instance of Visser’s Scheme, as (3.6) shows. However, this formula is derivable in the system iPV: it is easy to see that it follows from (3.6), using the fact that \( (\Box p_1 \lor \Box p_2 \lor \Box p_3) \) implies the formula \( ((\Box p_1 \rightarrow q) \rightarrow \Box p_1) \lor ((\Box p_1 \rightarrow q) \rightarrow \Box p_2) \lor ((\Box p_1 \rightarrow q) \rightarrow \Box p_3)) \) by propositional logic. Similar reasoning shows that the logic iPV is closed under substitution. However, in contrast to the other principles of iPH, this example shows that the collection of all instances of Visser’s Scheme is not closed under substitution.

Visser’s Scheme versus Visser’s Real Scheme

Although Visser’s Real Scheme and Visser’s Scheme are interderivable, it is not difficult to see that \( \{A\}(B) \) derives \( (A)(B) \), while in general the converse does not
hold. In this sense Visser's Real Scheme is more efficient than Visser's Scheme. Let us also illustrate this with one example.

Let \( A_1 = (p_1 \lor p_2) \) and \( A_2 = ((p_3 \lor p_4) \lor (p_5 \land p_6)) \), and consider the formula \( A = ((A_1 \rightarrow q) \rightarrow A_2) \). It is clear that

\[
\{A\}(A_1, A_2) = \left( \bigvee_{i=1}^{4} (A \rightarrow p_i) \lor ((A \rightarrow p_5) \land (A \rightarrow p_6)) \right).
\]

Thus by Proposition 3.2.2 it follows that

\[
\vdash_{iPV} A \triangleright \left( \bigvee_{i=1}^{4} (A \rightarrow p_i) \lor ((A \rightarrow p_5) \land (A \rightarrow p_6)) \right).
\]

However, while the derived formula is an instance (hence just one application) of Visser’s Real Scheme, this derivation in iPV uses many application of Visser’s Scheme. Namely, the application of Visser’s Scheme to \( A \) is

\[
A \triangleright ((A \rightarrow p_1 \lor p_2) \lor (A \rightarrow p_3 \lor p_4) \lor ((A \rightarrow p_5) \land (A \rightarrow p_6))).
\]

Is is clear that \( (\bigvee_{i=1}^{4} (A' \rightarrow p_i) \lor ((A' \rightarrow p_5) \land (A' \rightarrow p_6))) \) derives the formula \( ((A \rightarrow p_1 \lor p_2) \lor (A \rightarrow p_3 \lor p_4) \lor ((A \rightarrow p_5) \land (A \rightarrow p_6))) \), but not vice versa.

Note that if \( A \triangleright B \) is an instance of one of the schemes, then \( B \) derives \( A \), while in general the converse does not hold. Thus \( A \) can be a stronger formula than \( A \) (see the previous examples). For now, let us call a formula simple when either it is a propositional variable, a preservation or it is an implication for which either the antecedent is not a conjunct of implications or the consequent is a propositional variable, an implication or a preservation. Note that for simple formulas, the application of Visser’s Real Scheme or Visser’s Scheme does not lead to stronger formulas. We do not prove this fact, but the previous discussion indicates that if \( A \triangleright B \) is an instance of Visser’s Real Scheme, then every subformula of \( B \) that is not in the scope of an implication is simple. Therefore, the application of Visser’s Real Scheme does no longer lead to stronger formulas. This does not hold for Visser’s Scheme, as the example above shows. Thus in this sense Visser’s Real Scheme is more efficient than Visser’s Scheme. However, as mentioned before, for the modal study of the logic given by the scheme, we prefer to work with Vissers Scheme instead of Visser’s Real Scheme.

### 3.3 Preservativity versus provability

In this section we explain that the logic iP\( \mathcal{H} \) is contained in the system iP\( \mathcal{H} \), i.e. the principles of the provability logic of HA discussed in Section 2.5 are derivable in iP\( \mathcal{H} \). Then we show that the converse does not hold: iP\( \mathcal{H} \) derives principles in the language of provability which are not captured by the system iP\( \mathcal{H} \). These two facts show that iP\( \mathcal{H} \) is properly contained in the \( \mathcal{L}_\Box \)-part of iP\( \mathcal{H} \). It would be
interesting to know if one can obtain a decent axiomatization of the $L_\Box$-part of $i\Phi$. Although we did not find such an axiomatization yet, we conjecture that it exists.

First of all, the Necessitation Rule $(A/\Box A)$ is an admissible rule for $i\Phi$: If $\vdash A$ then $\vdash (T \rightarrow A)$. Hence by the Preservation Rule $\vdash T \rightarrow A$, which is $\vdash \Box A$. Lemma 3.1.1 shows that the axioms $K$, $4$ and $L$ belong to $i\Phi$. Leivant's Principle can be derived as follows.

\[ \vdash_{i\Phi} A \rightarrow A \wedge B \rightarrow B \]
\[ A \vdash (A \vee \Box B) \wedge B \vdash (A \vee \Box B) \quad \text{(Lemma 3.1.4)}(P1) \]
\[ (A \vee B) \vdash (A \vee \Box B) \quad \text{(Dp)} \]
\[ \Box (A \vee B) \rightarrow \Box (A \vee \Box B). \quad \text{(Lemma 3.1.1)} \]

Finally, we have to see that the Formalized Markov Scheme belongs to $i\Phi$. It is easy to see that $\neg \neg \vee_i \Box B_i \vdash (\vee_i \Box B_i)$ is derivable from Visser's Scheme:

\[ \vdash_{i\Phi} \neg \neg \vee_i^n \Box B_i \iff \neg (\wedge_i^n \neg B_i) \]
\[ \neg (\wedge_i^n \neg B_i) \vdash (\wedge_i^n \neg \Box B_i)(\bot, \Box B_1, \ldots, \Box B_n) \]
\[ (\wedge_i^n \neg \Box B_i)(\bot, \Box B_1, \ldots, \Box B_n) = \vee_i^n \Box B_i. \]

By Montagna's Principle we then have $(\Box A \rightarrow \neg \neg \vee_i \Box B_i) \vdash (\Box A \rightarrow \vee_i \Box B_i)$. Since $\neg (\Box A \rightarrow \vee_i \Box B_i)$ implies, $(\Box A \rightarrow \neg \neg \vee_i \Box B_i)$, this leads to the Formalized Markov Scheme $\neg (\Box A \rightarrow \vee_i \Box B_i) \vdash (\Box A \rightarrow \vee_i \Box B_i)$.

We show that $i\Phi$ is not conservative over $iH$. Note that for all axioms of $iH$ of the form $(\Box A \rightarrow \Box B)$, $i\Phi$ derives $A \rightarrow B$. For example, $(\Box A \rightarrow A) \vdash A$ and $(\neg \neg \Box B) \vdash \Box B$ belong to $i\Phi$. Using the Disjunctive Principle it follows that $i\Phi$ derives $((\Box A \rightarrow A) \vee \neg \neg \Box B) \vdash (A \vee \Box B)$ as well. Therefore, by Lemma 3.1.1 also $\Box((\Box A \rightarrow A) \vee \neg \neg \Box B) \rightarrow (\Box A \vee \Box B)$ is derivable in $i\Phi$. In Section 5.4 we show that this formula does not belong to $iH$.

The observation above has some interesting consequences. For example it shows that

\[ \vdash_{i\Phi} (\Box A \rightarrow A) \vee (\Box B \rightarrow B) \vdash (A \vee B). \]

And hence $\Box((\Box A \rightarrow A) \vee (\Box B \rightarrow B)) \rightarrow (\Box A \vee B)$ holds for HA, a principle which does not hold for PA.

### 3.4 Preliminaries

In this section we introduce a semantics for the preservativity and provability operators, and we define the canonical model and the construction method. These
are all fairly standard definitions except for the way in which the operator $\triangleright$ is interpreted in models. This semantics for $\triangleright$ is an idea from Visser. We also define the ‘new’ notion of an extendible property. In the proofs that this or that logic is canonical we need extensions of given sets of formulas. These extensions are all special instances of a ‘general’ principle of extension, which gave rise to the definition of an extendible property. First we introduce all these notions for preservativity logic. Most definitions are similar for provability logic. The ones that do differ are discussed in Section 3.4.6.

### 3.4.1 Definitions

The language $\mathcal{L}_{\triangleright}$ of preservativity logic is that of propositional logic extended with one binary modal operator, $\triangleright$. We assume $\bot$ (falsum) and $\top$ (true) to be present as primitive symbols in our propositional language. Recall that $\square A$ is defined as $\top \triangleright A$. A formula of the form $A \triangleright B$ is called a preservation and a formula of the form $\square A$ is called a boxed formula. We adhere to some reading conventions and omit parentheses when possible. The negation binds stronger than $\triangleright$ which binds stronger than $\wedge$ and $\vee$, which in turn bind stronger than $\rightarrow$. We use a ‘sequent-calculus’ abbreviation: $\Gamma \triangleright \Delta$ is short for $\bigwedge \Gamma \triangleright \bigvee \Delta$.

A logic is a theory closed under substitution. We call the logic in $\mathcal{L}_{\triangleright}$ which has as axioms all tautologies of intuitionistic propositional logic $\text{IPC}$ and the principles $P1$, $P2$ (and $Dp$) and as rules Modus Ponens and the Preservation Rule (Section 2.2) the arithmetical (semantical) base preservativity logic and denote it with $\text{iP}^-$ ($\text{iP}$). Following the notation of (Chagrov and Zakharyaschev 1997) we define $\text{iP}(A \oplus B)$ to be the preservativity logic consisting of the axioms of $\text{iP}$ plus $A$ and $B$, and the Preservation Rule and Modus Ponens. When $X$ denotes the infinite set of principles $A_1, A_2, \ldots$, we also write $\text{iP}X$ for $\text{iP}(A_1 \oplus A_2 \oplus \ldots)$. When $Xp$ is one of the principles of the preservativity logic given above we write $\text{iP}X$ for $\text{iP}Xp$.

We write $\vdash_{\text{iT}} A$ when $A$ is derivable in $\text{iT}$. We write $\Gamma \vdash_{\text{iT}} A$ when there is a derivation of $A$ in $\text{iT}$ from $\Gamma$ without use of the Preservation Rule, in other words, when $A$ is derivable by Modus Ponens from theorems of $\text{iT}$ and formulae in $\Gamma$.

The name ‘semantical base preservativity logic’ for $\text{iP}$ arises from the fact that it is sound and complete with respect to the frame semantics defined in Section 3.4.2. Thus, semantically seen, it is a base preservativity logic. On the other hand, the only axioms of $\text{iP}$ for which it is trivial to see that $\text{HA}$ derives all their arithmetical realizations are $\text{Taut}$, $P1$ and $P2$, and this accounts for the name ‘arithmetical base preservativity logic’ for $\text{iP}^-$. 

### 3.4.2 Semantics

A possible semantics for preservativity logic can be produced via frames: we just add one extra clause for the interpretation of $\triangleright$. The frames we use occur already in the literature (Section 2.6). The semantics for $\triangleright$ came from Visser.
3.4. Preliminaries

First some notation. When $R$ and $S$ are two binary relations, $(R;S)$ is the relation defined via $w(R;S)u \equiv \exists v(wRvSu)$.

A frame is a triple $F = (W, \preceq, R)$, where $W$ is a nonempty set (the set of nodes), $\preceq$ is a partial ordering on $W$ (the intuitionistic relation) and $R$ a binary relation on $W$ (the modal relation) such that $(\preceq;R) \subseteq R$.

A model is a quadruple $M = (W, \preceq, R, V)$, where $(W, \preceq, R)$ is a frame and $V$ a valuation relation on pairs consisting of nodes and propositional variables. We demand that $V$ is persistent, i.e.

\[(\text{persistence}) \quad \text{if } w \preceq v \text{ and } vVp, \text{ then } vVp.\]

We inductively define what it means for a formula $A$ to be forced (or valid) at a node $w$ of a model $M$ ($M, w \vdash A$):

\[
\begin{align*}
M, w \vdash p & \quad \equiv_{\text{def}} \quad wVp \\
M, w \vdash A \land B & \quad \equiv_{\text{def}} \quad M, w \vdash A \text{ and } M, w \vdash B \\
M, w \vdash A \lor B & \quad \equiv_{\text{def}} \quad M, w \vdash A \text{ or } M, w \vdash B \\
M, w \vdash A \to B & \quad \equiv_{\text{def}} \quad \forall v \preceq w (M, v \vdash A \implies M, v \vdash B) \\
M, w \vdash A \bigtriangledown B & \quad \equiv_{\text{def}} \quad \forall v \ (\text{if } wRv \text{ and } M, v \vdash A \text{ then } M, v \vdash B) \\
M, w \vdash \square A & \quad \equiv_{\text{def}} \quad \forall v \ (\text{if } wRv \text{ then } M, v \vdash A).
\end{align*}
\]

Note that the definition of forcing for $\square A$ agrees with the fact that $\square A$ is defined as $\top \to A$, and that $\square A$ gets the standard interpretation on frames. When $M$ is clear from the context we write $w \vdash A$ instead of $M, w \vdash A$. The formula $A$ is valid or forced in $M$, notation $M \models A$, if $A$ is forced in all nodes in $M$. The formula $A$ is valid in a frame $F$, notation $F \models A$, if $A$ is valid in all models with underlying frame $F$.

Note that $w \vdash A$ and $w \preceq v$ implies $v \vdash A$, and that $w \vdash \square A$ and $wRv$ implies $v \vdash A$.

A node $v$ in a frame is called a successor of $w$ if $wRv$, in which case $w$ is called a predecessor of $v$. We use an abbreviation for the relation $(R;\preceq)$:

\[\tilde{R} \equiv_{\text{def}} (R;\preceq).\]

For a relation $R$ we define $wR = \{v \mid wRv\}$. For a set $U$, we write $u \preceq U$ if for all $x \in U$, $u \preceq x$. We write \(\text{`}x \preceq y_1, \ldots, y_n\text{'}\) for \(\text{`}x \preceq y_1 \land x \preceq y_2 \land \ldots \land x \preceq y_n\text{'}\). Similarly for other relations. A node $v$ in a frame is above $w$ if $w \preceq v$. In this case $w$ is called below $v$. A node $x$ is called an (intuitionistic) top node if there is no element above it except the node itself. $\text{Top}(F)$ is the set of all top nodes in a frame $F$. We write $\text{Top}$ instead of $\text{Top}(F)$ if no confusion is possible. When $w$ is a node, $T(w)$ is the set of all top nodes above $w$. 

3.4.1. Remark. The condition $(\leq; R) \subseteq R$, included to guarantee persistence for formulas $A \rhd B$, may be weakened to

$$(\leq; R) \subseteq (R; \leq) \quad (w \preceq w' R v' \Rightarrow \exists u (w R u \preceq v')).$$

However we prefer to work with the simple condition where possible. For more discussion on this topic, see (Simpson 1993).

A property $P$ on frames corresponds to a set $T$ of formulas if for all frames $\mathcal{F}$: $\mathcal{F} \models T$ iff $F$ has property $P$. Note that in this case we have

if $\vdash_{iT} A$ then $A$ is valid on all frames with property $P$.

When a frame $\mathcal{F}$ has a property $P$ we say that $\mathcal{F}$ is a $P$-frame. We call $\mathcal{F}$ a $P_1 \ldots P_n$-frame when it has the properties $P_1 \ldots P_n$. If $C$ is a class of frames, a logic $iT$ is called complete with respect to $C$ if

for all $A$: $\vdash_{iT} A$ iff $A$ valid on all frames in $C$.

The logic $iT$ is called complete if $C$ is the class of frames to which $iT$ corresponds.

3.4.3 Canonicity

Canonical models are defined in a similar manner as in classical modal logic. To define the canonical $(X)$-model for a logic we have to introduce the notion of an $X$-saturated set. A set of formulas $X$ is called adequate if it is closed under subformulas and contains $\top$ and $\bot$. A set of formulas $\Gamma$ is called $X$-saturated with respect to a logic $T$ if it is a consistent subset of $X$ such that

- $\Gamma \vdash_T A$ implies $A \in \Gamma$, for all $A \in X$,
- $\Gamma \vdash_T A \lor B$ implies $A \in \Gamma$ or $B \in \Gamma$, for all $A \lor B \in X$.

If $X$ is the set of all formulas, an $X$-saturated set is just called saturated. It can be easily seen that for any (finite) adequate set $X$ and for any $A$ for which $\not\vdash A$, there is an (finite) $X$-saturated set $\Gamma$ such that $\Gamma \not\vdash A$. Note also that any $\Delta \subseteq X$ for which $\Delta \not\vdash A$, can be extended to an $X$-saturated $\Gamma$ such that $\Gamma \not\vdash A$.

For any logic $T$, for any adequate set $X$, the $T$-canonical $X$-model is the model $(W, \leq, R, V)$ defined as follows:

$W$ consists of the $X$-saturated sets (with respect to $\vdash_T$)

$w \preceq v \iff w \subseteq v$

$wRv \iff$ if $A_1, \ldots, A_n, B \in X, w \vdash_T A_1, \ldots, A_n \rhd B$ and $A_1, \ldots, A_n \in v$, then $B \in v$

$w \models p \iff p \in w$, for propositional variables $p \in X$. 
Recall that $A_1, \ldots, A_n \supset B$ is short for $(\land A_i) \supset B$ (Section 3.4.1). Note that in the definition of $R$ we take formulas $A \supset B$ into account which do not belong to $X$.

To see that this indeed defines a model, see the completeness proof for $iP$. This proof shows another fact we will often use, namely that for any canonical $X$-model:

$$\text{for all nodes } w, \text{ for all } A \in X : w \models A \text{ iff } A \in w.$$ 

When $X$ is the set of all formulas, we call the canonical $X$-model the \emph{canonical model of $T$}. We call a logic $iT$ \emph{canonical} if the canonical model has the frame property to which the logic corresponds. Note that canonical logics are always complete, namely with respect to the class of frames to which they correspond.

Note that in the $iT$-canonical frame in general $(R; \leq) \subseteq R$ does not hold. On the other hand, if we restrict our language to $\Box$ and the connectives, the canonical models do satisfy $(R; \leq) \subseteq R$, see Section 3.4.6. That $(R; \leq) \subseteq R$ is too strong a requirement in the context of preservativity logic follows from the fact that $A \supset B \rightarrow \Box(A \supset B)$ is valid on such frames. This principle is not in the preservativity logic of HA as was explained in Section 3.1.

### 3.4.4 Extendible properties

In this section we introduce a general construction to make certain extensions of sets of formulas. In many proofs to come we will extend certain sets of formulas to saturated sets with certain properties. It turns out that the way these extensions are made follow the same pattern. Therefore, we choose to define a general notion of extension which covers this.

Let $iT$ be a preservativity logic and $X$ an adequate set (Section 3.4.3). A property $*(\cdot)$ on sets of formulas such that we have both

- for all $A \in X$: if $*(x)$ and $x \vdash_{iT} A$, then $*(x \cup \{A\})$
- for all $(A \lor B) \in X$: if $*(x \cup \{A \lor B\})$, then $*(x \cup \{A\})$ or $*(x \cup \{B\})$,

is called an $iT$-\emph{extendible property} (w.r.t. $X$). If in addition it holds that

- for all $A \in X$: if $*(x)$ and $y \vdash_{iT} x \supset A$, then $*(x \cup \{A\})$

then it is called an $iT$-\emph{extendible $y$-successor property}. For a property $*$ such that $*(\Gamma)$ holds, the $*\text{-extension of } \Gamma$ is the union $x = \bigcup x_i$ of sets $x_i$ which are constructed as follows. Given an enumeration $B_0, B_1, \ldots$ of all formulas in $X$, in
Chapter 3. Tools and preliminaries

\[ \forall x, \{ \text{Xi} \} \subseteq \text{U} \{ \text{Xi} \} \]

if \( f \) is \( \text{E} \), no disjunction \( \text{E} \) = \text{CVD}, \( f \) = \text{DC}.

whenever a formula occurs infinitely often, we define
\[ x_0 = \Gamma \]

3.4.2. Remark. Note that for an \( \text{I} \)-extendible \( \text{y}-\text{successor} \) property, the first requirement is redundant, because it follows from the third one. Namely, if \( x \rightarrow A \) holds we have \( I \rightarrow (x \rightarrow A) \), and hence by Preservation Rule \( x \rightarrow A \).

In the completeness proofs in the next chapters we often use extendible properties in the following way. Given a set \( A \) with a certain property, we want to extend it to a saturated set with this property. There are two particular properties which often occur in this setting. The following lemma shows that these properties are extendible to successor properties.

3.4.3. Lemma. For any logic \( \text{I} \) containing \( \text{I} \), for any formula \( C \) and for all nodes \( w, v \) in the \( \text{I} \)-canonical model, the following two properties are extendible to \( \text{I} \)-successor properties:

\[ \text{w} \rightarrow \text{E} \]

\[ \text{w} \rightarrow \text{C} \]

\[ \text{E} \subseteq \text{D} \]

3.4.4. Proof. We write \( \text{h} \) for \( \text{h} \). First we consider the property \( \text{w} \rightarrow \text{E} \). We have to show that for all \( x \in A \):

\[ \text{if w} \rightarrow \text{C} \text{ and z} \rightarrow A, \text{ then w} \rightarrow \text{z \rightarrow E} \]

or

\[ \text{if w} \rightarrow \text{A} \text{ and w} \rightarrow \text{z} \rightarrow A, \text{ then w} \rightarrow \text{z \rightarrow A} \]

Observe that \( x \rightarrow \text{Gamma} \) is \( X \)-saturated. Thus, \( x \) is a node in the canonical \( X \)-model, and if \( \Gamma \) is also a node in the canonical \( X \)-model, then also \( x \rightarrow \text{Gamma} \) is an \( \text{I} \)-extendible \( \text{y}-\text{successor} \) property.

3.4.5. Remark. Note that for an \( \text{I} \)-extendible \( \text{y}-\text{successor} \) property, the first requirement is redundant, because it follows from the third one. Namely, if \( x \rightarrow A \) holds we have \( I \rightarrow (x \rightarrow A) \), and hence by Preservation Rule \( x \rightarrow A \). Thus clearly \( w \rightarrow \text{I} \rightarrow \text{A} \).

In the completeness proofs in the next chapters we often use extendible properties in the following way. Given a set \( A \) with a certain property, we want to extend it to a saturated set with this property. There are two particular properties which often occur in this setting. The following lemma shows that these properties are extendible to successor properties.

3.4.6. Lemma. For any logic \( \text{I} \) containing \( \text{I} \), for any formula \( C \) and for all nodes \( w, v \) in the \( \text{I} \)-canonical model, the following two properties are extendible to successor properties:

\[ \text{w} \rightarrow \text{E} \]

\[ \text{w} \rightarrow \text{C} \]

\[ \text{E} \subseteq \text{D} \]

3.4.7. Proof. We write \( \text{h} \) for \( \text{h} \). First we consider the property \( \text{w} \rightarrow \text{E} \). We have to show that for all \( x \in A \):

\[ \text{if w} \rightarrow \text{C} \text{ and z} \rightarrow A, \text{ then w} \rightarrow \text{z \rightarrow E} \]

or

\[ \text{if w} \rightarrow \text{A} \text{ and w} \rightarrow \text{z} \rightarrow A, \text{ then w} \rightarrow \text{z \rightarrow A} \]
Recall that we write \( x, A \triangleright C \) for \((\bigwedge x \land A) \triangleright C\). By Remark 3.4.2 we know that if the third requirement holds, so does the first. Therefore, it suffices to show that the last two requirements holds.

For the second requirement, assume \( w \vdash x, A \triangleright C \) and \( w \vdash x, B \triangleright C \). To show that \( \ast(\cdot) \) satisfies the second requirement we have to prove that \( w \vdash x, (A \lor B) \triangleright C \).

This follows immediately from \( Dp \).

For the third requirement assume \( w \vdash x \triangleright A \) and \( w \vdash x, A \triangleright C \). We show that \( w \vdash x \triangleright C \), and this will show that \( \ast(\cdot) \) satisfies the third requirement. By the Preservation Rule we have \( \vdash x \triangleright \bigwedge x \), which is short for \( \vdash \bigwedge x \triangleright \bigwedge x \). Therefore, we certainly have \( w \vdash x \triangleright \bigwedge x \). Thus by \textit{P} 2 we have \( w \vdash x \triangleright (\bigwedge x \land A) \). Together with \( w \vdash x, A \triangleright C \) and \( P1 \) this leads to \( w \vdash x \triangleright C \).

Consider the property \( \ast \). To show that \( \ast \) is an extendible \( w \)-successor property we have to prove that

\[
\text{for all } A \in X: \quad \text{if } \ast(x) \text{ and } x \vdash A, \text{ then}
\]

\[
(\text{for all } D: w \vdash x, A \triangleright D \text{ implies } D \in v)
\]

\[
\text{for all } (A \lor B) \in X: \quad \text{if } \ast\left(x \cup \{A \lor B\}\right), \text{ then}
\]

\[
(\text{for all } D: w \vdash x, A \triangleright D \text{ implies } D \in v) \quad \text{or}
\]

\[
(\text{for all } D: w \vdash x, B \triangleright D \text{ implies } D \in v)
\]

\[
\text{for all } A \in X: \quad \text{if } \ast(x) \text{ and } w \vdash x \triangleright A, \text{ then}
\]

\[
(\text{for all } D: w \vdash x, A \triangleright D \text{ implies } D \in v).
\]

By Remark 3.4.2, it suffices to show that the last two requirements holds.

For the second requirement, assume that neither \( \ast(x \cup \{A\}) \) nor \( \ast(x \cup \{B\}) \) holds. We prove that \( \ast(x \cup \{A \lor B\}) \) does not hold. By assumption there are formulas \( C \) and \( D \) such that \( C \not\in v \) and \( D \not\in v \), and both \( w \vdash x, A \triangleright C \) and \( w \vdash x, B \triangleright D \). Clearly, both \( C \) and \( D \) imply \( (C \lor D) \). Hence by Preservation Rule we have \( \vdash C \triangleright (C \lor D) \) and \( \vdash D \triangleright (C \lor D) \). Applying \( P1 \) gives \( w \vdash x, A \triangleright (C \lor D) \) and \( w \vdash x, B \triangleright (C \lor D) \). Thus by \( Dp \) we have \( w \vdash x, (A \lor B) \triangleright (C \lor D) \). If \( \ast(x \cup \{A \lor B\}) \) would hold, this would imply that \( (C \lor D) \in v \). Since \( v \) is a node in the canonical model it is a saturated set. Therefore, this would imply that \( C \in v \) or \( D \in v \), which contradicts our assumption.

We show that the third requirement holds. Assume that \( \ast(x) \) and \( w \vdash x \triangleright A \) hold, and that we have \( w \vdash x, A \triangleright D \), for some \( D \). We have to show that \( D \in v \). The same reasoning as above for \( \ast(\cdot) \), shows that we have \( w \vdash x \triangleright (\bigwedge x \land A) \). Therefore, \( w \vdash x, A \triangleright D \) implies \( w \vdash x \triangleright D \) by \( P1 \). The fact that \( \ast(x) \) holds, gives \( D \in v \).
3.4.5 The Construction Method

We define a method, the construction method, to obtain from a given model a new one. This method is similar to the construction method in classical model logic. The construction method is often used to obtain a completeness result with respect to some class of finite frames. Let \( \mathcal{M} = (W, \preceq, R, V) \) be some canonical model, let \( X \) be an adequate set for which \( A \rightarrow B \in X \) implies \( \Box B \in X \). The method allows us to construct for any \( w \in W \) a model \( \mathcal{M}' = (W', \preceq', R', V') \) the domain of which consists of (copies of) nodes in \( W \), which intuitively is the minimal set of nodes required to have \( w \) forcing the same formulae in \( X \) in the models \( \mathcal{M} \) and \( \mathcal{M}' \). We will restrict ourselves to a construction method for models that besides iPP also satisfy LP and MP.

The construction proceeds as follows. We choose step by step, starting with \( w \), a subset of \( W \) which will be the domain \( W' \) of our new model \( \mathcal{M}' \). Note that the elements of \( W \) are sets of formula. First, define

\[
\begin{align*}
\forall X \in W & : \{ A \rightarrow B \in X \mid A \rightarrow B \in w \}, \\
\forall X \in W & : \{ A \rightarrow B \in X \mid A \rightarrow B \not\in w \}.
\end{align*}
\]

Similarly for \( \rightarrow \). We omit the superscript \( X \) when possible. Let \( * \) denote the concatenation function on strings:

\[
\langle x_1, \ldots, x_n \rangle * \langle y_1, \ldots, y_m \rangle = \langle x_1, \ldots, x_n, y_1, \ldots, y_m \rangle.
\]

Put \( \alpha() = w \). Suppose \( v = \alpha_\gamma \) is defined. We choose elements \( \alpha_{\sigma \langle A \rightarrow B \rangle} \) and \( \alpha_{\sigma \langle A \rightarrow B \rangle} \) in \( W \), for all elements \( (A \rightarrow B) \in \sigma \not\gamma \), \( A \rightarrow B \in \sigma \not\gamma \).

The node \( \alpha_{\sigma \langle A \rightarrow B \rangle} \) is an element \( u \in W \) such that \( v \preceq u \), \( A \in u \) and \( B \not\in u \). Note that such elements can always be found. The node \( \alpha_{\sigma \langle A \rightarrow B \rangle} \) is an element \( u \in W \) such that \( vRu, A \in u \), \( B \not\in u \) and \( \Box B \in v \). Observe that \( u \) contains more boxed formulae than \( v \), for in the presence of LP, and hence of 4p and 4, \( vRu \) and \( \Box C \in v \) implies that \( \Box C \in u \). To prove that such a node \( u \) exists it suffices to show that in any canonical model for a logic containing iPLM, if \( A \rightarrow B \not\in v \) there exists a \( v \)-successor extension of \( \{A, \Box B\} \) omitting \( \{B\} \). Thus we have to see that \( v \vdash A, \Box B \rightarrow B \). Suppose not. Then we have, using LP and MP:

\[
\begin{align*}
v & \vdash A, \Box B \rightarrow B \\
& \Box B \rightarrow A \land \Box B \rightarrow (\Box B \rightarrow B) \\
& A \rightarrow (\Box B \rightarrow B) \\
& A \rightarrow B.
\end{align*}
\]

Define \( W' = \{ \sigma \mid \sigma \text{ is defined } \} \), and \( V \) via

\[
\sigma \models p \iff \alpha_\gamma \models p, \text{ for } p \in X.
\]
We define the intuitionistic and the modal relation such that

for all $A \in X$, for all $\sigma \in W'$: $\alpha_\sigma \vdash A$ iff $\sigma \vdash A$.

As the choice of the relations will differ from case to case we do not give any specific examples here besides the obvious one;

$$\sigma \preceq \tau \iff \alpha_\sigma \preceq \alpha_\tau$$
$$\sigma R^I \tau \iff \alpha_\sigma R^I \alpha_\tau.$$

It is not difficult to see that this choice gives a model with the desired property, be it not always on a frame with the desired properties.

3.4.4. Remark. It is easy to see that $W'$ is finite if $X$ is. First note that by construction, a node (saturated set) $\sigma \star (B \supset C)$ contains more boxed formulas (formulas of the form $\Box C$) that belong to $X$ than $\sigma$. A node $\sigma \star (B \rightarrow C)$ contains more implications that belong to $X$ than $\sigma$. Moreover, for a node $\tau = \sigma \star (B \rightarrow C)$ we have that $\alpha_\sigma \preceq \alpha_\tau$ holds in the canonical model, i.e. $\alpha_\sigma \subseteq \alpha_\tau$. Clearly, all the implications that have to be treated, i.e. all implications for which we possibly have to add a new node in the construction, belong to $X$. And similarly for boxed formulas and preservations. Therefore, in going from $\sigma$ to $\sigma \star (B \supset C)$ or $\sigma \star (B \rightarrow C)$ either the number of boxed formulas that have to be treated decreases, or it stays the same and the number of implications that have to be treated decreases. Finally, if there are no more boxed formulas to be treated this means that for all $\Box B \in X$, it holds that $\Box B \in \alpha_\sigma$. Hence for all $B \supset C \in X$, we have $\Box C \in \alpha_\sigma$ and thus $B \supset C \in \alpha_\sigma$. Therefore, if there are no more boxed formulas to be treated there are no formulas of the form $B \supset C$ to be treated either. Since the preservations and implications that belong to $X$ are the only formulas that have to be treated in the construction method, this shows that the method is finite if $X$ is.

3.4.6 The language of provability logic

The language of provability logic $L_\Box$ is that of propositional logic extended with one modal operator $\Box$. We write

$$\Box A \equiv_{df} A \land \Box A.$$ 

The definition of $w_\square$ is similar to $w_P$.

For any principles $A$ and $B$, $iK(A \oplus B)$ is the logic in $L_\Box$ consisting of all formulas provable in intuitionistic propositional logic IPC and the axioms $K$ plus $A$ and $B$, and the rules Modus Ponens and Necessitation $(C/\Box C)$. As in classical provability logic, we write $iT$ for $iK$, for any set of principles $T$. We write $\Gamma \vdash_{iT} A$ when $A$ is derivable in $iT$. We write $\Gamma \vdash_{iT} A$ when there is a derivation of $A$ in $iT$ from $\Gamma$ without use of Necessitation, in other words, when $A$ is derivable by Modus Ponens from theorems of $iT$ and formulae in $\Gamma$. 

A (non)boxed formula is a formula (not) of the form $\square A$.

The definition of a frame, a model and the notion of correspondence are inherited from preservativity logic, by reading $T \triangleright A$ for $\square A$. As observed before, in that way $\square A$ gets the standard interpretation on frames.

**Canonicity in provability logic**

It is convenient to change the definition of a canonical model (Section 3.4.3) slightly in the context of provability logic. For any logic $T$ in $\mathcal{L}_{\square}$ and for any adequate set $X$, the *$T$-canonical $X$-model* is the model $(W, \preceq, R, V)$ defined as follows:

- $W$ consists of the $X$-saturated sets (with respect to $\vdash_T$)
- $w \preceq v \iff w \subseteq v$
- $wRv \iff$ if $\square A \in w$ then $A \in v$
- $w \vdash p \iff p \in w$, for propositional variables $p \in X$.

Given this definition, the definition of canonicity is similar to the one in preservativity logic. The difference between this definition on canonical model and the one in the context of preservativity logic lies in the definition of $R$, which for the latter would read $wRv$ iff for all $A \in X$, if $w \vdash \square A$ then $A \in v$.

**Brilliant frames**

Recall that in a frame we always have $(\preceq; R) \subseteq R$. A frame is called *brilliant* if in addition it holds that

$$\textit{(brilliant)} \quad \tilde{R} \subseteq R$$

where $\tilde{R}$ is defined as $(R; \preceq)$ (Section 3.4.2). Note that in $\mathcal{L}_{\square}$, canonical models have brilliant frames. In $\mathcal{L}_{\triangleright}$ they do not have this property. For example, $A \triangleright B \rightarrow \square (A \rightarrow B)$ is valid on these frames, a principle which is not arithmetically valid (see Section 3.1). However, we will see that if we restrict ourselves to $\mathcal{L}_{\square}$ all provability principles considered are complete with respect to some class of brilliant frames, even though they are sometimes also complete with respect to some nice class of non-brilliant frames.

**Extendible properties in provability logic**

The definition of an extendible property (Section 3.4.4) does not change in the context of $\mathcal{L}_{\square}$.

**3.4.5. Remark.** Let $*$ be a extendible property w.r.t. an adequate set $X$. Note that if $x$ is the $*$-extension of a set which contains $\{A \mid \square A \in y\}$, then $x$ is a node in the canonical $X$-model and in this model $yRx$ holds.