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 2

Concepts

In this chapter we introduce the notions studied in part I of the thesis. Sections 2.1, 2.2 and 2.5 explain what provability, preservativity, and intuitionistic provability logic are. We introduce preservativity and intuitionistic provability logic in more detail than classical provability logic, for which there are many nice overview articles and books: (Boolos 1979) (Smoryński 1985) (Boolos 1993) (Japaridze and de Jongh 1997) (Visser 1997). Section 2.6 briefly summarizes the literature on intuitionistic modal logic and explains how in the case of provability logic the presented results deviate from the regular literature. Section 2.8 gives an overview of the next chapters of part I.

2.1 Provability logic

In Chapter 1 we explained the idea behind provability logic. There we saw that the provability logic of an arithmetical theory $T$ consists of all the propositional schemes that $T$ can prove about its provability predicate. In this section we give the formal definition of provability logic. Recall that $\Box_T$ denotes the provability predicate of $T$ and that a sentence $\Box_T\varphi$ expresses the statement that $\varphi$ is provable in $T$.

Let $\mathcal{L}_\Box$ be the language of propositional logic extended with one modal operator $\Box$. The formulas in $\mathcal{L}_\Box$ are called modal formulas. Let $T$ be an arithmetical theory that is strong enough to allow the formalization of its provability predicate. An arithmetical realization of $\mathcal{L}_\Box$ into the language of $T$ is a mapping $*$ from the formulas of $\mathcal{L}_\Box$ to sentences in the language of $T$ that commutes with the propositional connectives and such that $(\Box A)^* = \Box_T(\varphi A^*)$. The provability logic of $T$ is the set of modal formulas $A$ such that $T$ proves $A^*$ for any arithmetical realization $*$, i.e. the set $\{A \mid \forall^* T \vdash A^*\}$. The truth provability logic of $T$ is the

---

$^1$We will not discuss the minimal requirements that such a theory should satisfy, they can be found in (Smoryński 1985) or (Hájek and Pudlák 1991).
set of modal formulas $A$ such that $A^*$ is valid in the standard model $\mathbb{N}$ for any arithmetical realization $\ast$, i.e. the set $\{A \mid \forall^* \mathbb{N} \models A^*\}$.

Note that in general the provability logic of a theory $T$ may depend on $T$ as well as on the chosen formalization of the proof predicate $Proof_T$. We will be a bit ambiguous in this respect. When talking about 'the provability logic' of a certain theory, we will always assume that a not-to-unusual proof predicate is fixed in advance.

The famous article by Solovay (1976) may well be seen as the starting point of provability logic. In this paper Solovay proves that the the provability logic of Peano Arithmetic PA is the logic now known as $L$ or GL, consisting of the principles $K$, $4$ and $L$ (Section 2.5), the tautologies of classical propositional logic and the rules Necessitation ($A/\Box A$) and Modus Ponens. Moreover, the proof gives a method to construct for any formula $A$ which is not a principle of the provability logic of PA an actual counterexample $A^*$, that is, a realization such that PA does not prove $A^*$. The way $A^*$ is obtained employs a modal completeness result for GL. First, it is shown that GL is complete with respect to the class of finite, transitive, conversely well-founded Kripke models. And second, it is shown that for every such Kripke model $\mathcal{K}$ there exists a realization $\ast$ such that

$$\text{for all nodes } k \text{ of } \mathcal{K} \text{ (if } \mathcal{K}, k \not\models A, \text{ then } \neg A^* \text{ is consistent with PA ).}$$

This shows the usefulness of a semantical characterization of provability logics.

As mentioned before, provability logics of classical theories are well-investigated. One remarkable thing is their stability; many theories, like for example PA and ZF, have the same provability logic, GL (Solovay 1976)(Visser 1984)(Beklemishev 1990). Although for intuitionistic theories we know much less, we do know that there is in general no stability in going from a classical theory to its intuitionistic counterpart, as can be seen in the comparison of HA and PA. This will be explained in Section 2.5 on intuitionistic provability logic.

It is clear that the idea behind provability logic can be applied to any predicate which can be encoded in arithmetical theories. We will see examples of this in the next section.

### 2.2 Preservativity logic

In this section we introduce the notion of $\Sigma_1$-preservativity which is an extension of provability. This notion was invented by Visser (1994) and arose from the study of the admissible rules of Heyting Arithmetic HA, the constructive theory of the natural numbers (a definition of HA can be found in (Troelstra and van Dalen 1988)). It turns out that many principles of the provability logic of Heyting Arithmetic have an elegant formulation in this setting. Therefore, the questions in provability logic can better be studied in the context of preservativity.
2.2. Preservativity logic

The definition of preservativity

Let $\Sigma_i$ and $\Pi_i$ denote the well-known levels of the arithmetical hierarchy (Hájek and Pudlák 1991). For an arithmetical theory $T$ and sentences $\varphi$ and $\psi$ in the language of $T$, $\varphi$ is said to $\Sigma_1$-preserve $\psi$ with respect to $T$, if for all $\Sigma_1$-sentences $\theta$ it holds that $T \vdash (\theta \rightarrow \varphi)$ implies $T \vdash (\theta \rightarrow \psi)$. We denote this with $\varphi \triangleright_T \psi$. Since we will not consider any other forms of preservativity than $\Sigma_1$-preservativity we will, as in the title, always refer to preservativity instead.

On the modal side the notion of preservativity gives rise to a modal language $L_\triangleright$ with one binary modal operator, $\triangleright$. Analogous to provability logic the preservativity logic of $T$ is defined as the collection of $L_\triangleright$-formulas $A$ such that $T \vdash A^*$ for any arithmetical realization $\ast$. In this context the definition of an ‘arithmetical realization’ is extended to cover formulas in which the preservativity symbol $\triangleright$ occurs: an arithmetical realization $\ast$ is a mapping from $L_\triangleright$-formulas to arithmetical formulas which commutes with the connectives and such that $(A \triangleright B)^* = \text{Pres}_T(A^*, B^*)$, where $\text{Pres}_T(x, y)$ is a formula in the language of $T$ that is the formalized version of the statement $A \triangleright_T B$. Like in the case for $L_\Box$, the formulas in $L_\triangleright$ are called modal formulas.

Clearly, preservativity is an extension of provability because we have

$$\Box_T \varphi \text{ iff } T \triangleright_T \varphi.$$ 

In Section 2.5 we will return to this relation with provability logic.

For classical theories $T$ the notion of preservativity is equivalent to the notion of $\Pi_1$-conservativity: we have that $\varphi$ $\Sigma_1$-preserves $\psi$ if and only if $\neg \varphi$ is $\Pi_1$-conservative over $\neg \psi$. For many classical theories, for example PA, the notion of $\Pi_1$-conservativity is again equivalent to the well-investigated notion of interpretability. Therefore, for these theories the preservativity logic is known, although the notion is not studied directly but only via the equivalence with interpretability. In Section 2.4 we will discuss the connection with interpretability logic in more detail.

For constructive theories like HA, the situation is completely different. In the next section we will explain how in this setting the notion of preservativity arises in a natural way from the admissible rules and that the admissible rules play a prominent role in the provability and preservativity logic of HA. Moreover, we will see that the notion of preservativity seems to give the right view on questions in provability logic of constructive theories.

On the classical side, the first theory studied in the context of provability logic was Peano Arithmetic PA (Hájek and Pudlák 1991), the well-known theory of the natural numbers, for it is strong enough to allow the formalization of provability notions in an easy way. For the same reason the questions in intuitionistic provability logic focus on Heyting Arithmetic HA (Troelstra and van Dalen 1988), the intuitionistic counterpart of PA. In this thesis we will only consider preservativity with respect to HA.
2.3 Heyting Arithmetic

An (c.e.) axiomatization of the preservativity logic (or the provability logic) of HA is not known. However, Visser (1994) has given some principles of the preservativity logic of HA, which capture all principles of the provability logic of HA known before that time. In the following years the meaning of these principles became clear (Visser 1999)(Iemhoff 2000b) (Chapter 7). These insights in the system led us to the conjecture that it is the preservativity logic of HA. In this section we introduce the system, discuss its meaning and explain why we conjecture it to be the preservativity logic of HA.

To state the principles of the preservativity logic of HA known so far, we need the following notation. For formulas $A, B_1, \ldots, B_n$, the formula $(A)(B_1, \ldots, B_n)$ is inductively defined to be

$$(A)(B, C_1, \ldots, C_n) \equiv_{\text{def}} (A)(B) \lor (A)(C_1, \ldots, C_n)$$

$$(A)(\bot) \equiv_{\text{def}} \bot$$

$$(A)(B \land B') \equiv_{\text{def}} (A)(B) \land (A)(B')$$

$$(A)(\Box B) \equiv_{\text{def}} \Box B$$

$$(A)(B) \equiv_{\text{def}} (A \rightarrow B)$$

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

Note that we have $(A)(C_1, \ldots, C_n) = (A)(C_1) \lor \ldots \lor (A)(C_n)$, and that $(A)(\top) = (A \rightarrow \top)$, hence $(A)(\top) \iff \top$.

The expression $\cdot(\cdot)$ is an abbreviation and not an operator, because applying it to equivalent formulas does not give equivalent results. For example, $\Box p$ is equivalent to $(\top \rightarrow \Box p)$, but $(A)(\top \rightarrow \Box p) = (A \rightarrow (\top \rightarrow \Box p))$ and $(A)(\Box p) = \Box p$. Hence the formulas $(A)(\top \rightarrow \Box p)$ and $(A)(\Box p)$ are in general not equivalent.

In (Visser 1994) the following principles of the preservativity logic of HA known so far are given. In fact, we give here a slightly different axiomatization than the one used by Visser. In Chapter 3 we will see that Visser’s system is equivalent to the one introduced here. We denote intuitionistic propositional logic with IPC (Troelstra and van Dalen 1988). Recall that $\varphi$ is provable if and only if $\top$ preserves $\varphi$. This accounts for the definition of $\Box$ in the system.
Principles of the preservativity logic of HA

\[ \Box A \equiv_{def} \top \vdash A \]

Taut all tautologies of IPC

\[ P1 \quad A \vdash B \land B \vdash C \rightarrow A \vdash C \]

\[ P2 \quad C \vdash A \land C \vdash B \rightarrow C \vdash (A \land B) \]

\[ Dp \quad A \vdash C \land B \vdash C \rightarrow (A \lor B) \vdash C \quad \text{(Disjunctive Principle)} \]

\[ 4p \quad A \vdash \Box A \]

\[ Lp \quad (\Box A \rightarrow A) \vdash A \quad \text{(Löb’s Preservativity Principle)} \]

\[ Mp \quad A \vdash B \rightarrow (\Box C \rightarrow A) \vdash (\Box C \rightarrow B) \quad \text{(Montagna’s Principle)} \]

\[ Vp_n \quad (\land_{i=1}^n A_i \rightarrow B_i) \vdash A_{n+1} \lor A_{n+2} \vdash (\land_{i=1}^n A_i \rightarrow B_i)(A_1, \ldots, A_{n+2}) \quad \text{(Visser’s Principles)} \]

\[ Vp \quad Vp_1, Vp_2, Vp_3, \ldots \quad \text{(Visser’s Scheme)} \]

We use the name iP H for the logic given by these principles and the rules Modus Ponens and the

Preservation Rule \quad \text{if } \vdash (A \rightarrow B) \text{ then } \vdash A \vdash B.\]

In Sections 2.4 and 2.5 we discuss the relation between the logic iP H and interpretability and provability logic. In Section 2.4 we will see that all principles except the Disjunctive Principle hold for PA as well. In Section 2.7 we repeat the proofs by Visser (1994) that these principles and rules belong to the preservativity logic of HA. Visser’s Scheme is a special and complicated scheme. In Section 3.2 we elaborate on the technical details of this scheme and show that our formulation of the scheme is equivalent to the one used by Visser (1994).

Here we discuss the meaning of the given principles. We will see that these principles form a natural fragment of the preservativity logic of HA. Namely, each of them corresponds to either a principle of the provability logic of PA or to one of the following characteristic properties of HA: its propositional admissible rules, Markov’s Rule and the Disjunction Property.

The definition of \( \Box \) and the first two principles are easily seen to be principles of the preservativity logic of HA. The principles 4p and Lp resemble the two characteristic axioms for the provability logic of PA, which are

\[ 4 \quad \Box A \rightarrow \Box \Box A \]

\[ L \quad \Box (\Box A \rightarrow A) \rightarrow \Box A. \]
Since $A \rightarrow B$ implies ($\Box A \rightarrow \Box B$) in the system (Section 3.1), the principles 4$p$ and $Lp$ imply their provability counterparts 4 and $L$. The principle 4 is derivable from $L$, but usually it is still included in the axioms. We will see that in the same way 4$p$ is derivable from $Lp$ (Section 4.3). The principle $Mp$ is baptized after its classical counterpart in interpretability logic, which is discussed below. It is easy to see that it belongs to the preservativity logic of HA, using the fact that the arithmetical realization of a formula $\Box C$ is always $\Sigma_1$ (Section 2.7).

The Disjunctive Principle and the Disjunction Property

The Disjunctive Principle $Dp$ is related to the Disjunction Property of HA, which reads

\[(\text{Disjunction Property}) \quad \text{if } HA \vdash \varphi \lor \psi, \then HA \vdash \varphi \lor HA \vdash \psi.\]

Friedman (1975) proved that HA does not prove its Disjunction Property, i.e. HA does not derive the true formula $\Box (\varphi \lor \psi) \rightarrow (\Box \varphi \lor \Box \psi)$. Leivant (1975) showed that HA does prove the weaker version

\[HA \vdash \Box (\varphi \lor \psi) \rightarrow \Box (\varphi \lor \Box \psi).\]

Hence the so-called Leivant Principle $\Box (A \lor B) \rightarrow \Box (A \lor \Box B)$ is part of the provability logic of HA. In the preservativity logic of HA this principle occurs as a consequence of the two principles 4$p$ and $Dp$. Note that the fact that $Dp$ and 4$p$ are in the preservativity logic of HA imply the following strengthening of Leivant’s Principle:

\[HA \vdash (\varphi \lor \psi) \Rightarrow (\varphi \lor \Box \psi).\]

The arithmetical validity of the Disjunctive Principle was shown by Visser (1994) and will be treated in Section 2.7.

Visser’s Scheme and the admissible rules

The scheme $Vp$ is called after A. Visser who proved its arithmetical validity (Visser 1994). Note that it is not a principle but a collection of infinitely many principles. They describe (some) admissible rules of HA. For propositional formulas $A, B$ we say that the rule $A/B$ is a propositional admissible rule of HA if $HA \vdash \sigma A$ implies $HA \vdash \sigma B$, for all substitutions $\sigma$ which replace the propositional variables by arithmetical formulas. Observe that if ($\Box A \rightarrow \Box B$) is in the provability logic of HA this implies that $A/B$ is an admissible rule of HA. Since $A \rightarrow B$ implies ($\Box A \rightarrow \Box B$), it follows that if $A \rightarrow B$ is in the preservativity logic of HA, then $A/B$ is an admissible rule for HA. The two most meaningful instances of $Vp$ describe the propositional admissible rules and Markov’s Rule for HA. We will discuss them briefly.

If one restricts Visser’s Scheme to pure propositional formulas, i.e. without $\Box$ or $\Rightarrow$, it characterizes the propositional admissible rules of HA, as will be proved in
Chapter 7. There we will see that if we let AR be the logic given by the principles Taut, P1, P2, Dp, Vp and the Preservation Rule, then we have

for propositional formulas $A, B$:

$A \vdash B$ is a propositional admissible rule of HA if $AR \vdash A \rightarrow B$.

This will be explained in more detail in Section 2.3.1. Here we consider one example. It is well-known that

$$\neg A \rightarrow B \vee C / (\neg A \rightarrow B) \vee (\neg A \rightarrow C)$$

is an admissible rule of HA (Harrop 1960). We show how we can derive the corresponding statement

$$(\neg A \rightarrow B \vee C) \vdash ((\neg A \rightarrow B) \vee (\neg A \rightarrow C))$$

in the system AR:

$$\vdash_{AR} (\neg A \rightarrow B \vee C) \vdash (\neg A)(A, B, C) \quad (Vp) \quad (1)$$

$$(\neg A)(A, B, C) \leftrightarrow (\neg A \rightarrow A) \vee (\neg A \rightarrow B) \vee (\neg A \rightarrow C) \quad (2)$$

$$\neg A \rightarrow A \quad (Taut) \quad (3)$$

$$\neg A \rightarrow (\neg A \rightarrow B) \quad (Taut) \quad (4)$$

$$\neg A \rightarrow A \vee (\neg A \rightarrow B) \vee (\neg A \rightarrow C) \rightarrow$$

$$(\neg A \rightarrow B) \vee (\neg A \rightarrow C) \quad (3)(4)(Taut) \quad (5)$$

$$(\neg A \rightarrow A) \vee (\neg A \rightarrow B) \vee (\neg A \rightarrow C) \vdash$$

$$\neg A \rightarrow B \vee (\neg A \rightarrow C) \quad (5)(PreservationRule) \quad (6)$$

$$\neg A \rightarrow B \vee (\neg A \rightarrow C) \vdash ((\neg A \rightarrow B) \vee (\neg A \rightarrow C)) \quad (1)(2)(6)(P1)$$

This shows that the admissible rule given above is captured by Visser's Scheme. Markov's Rule, a well-known rule for HA, reads

$(Markov's Rule)$ for all $\varphi \in \Pi_2$: if $HA \vdash \neg \neg \varphi$, then $HA \vdash \varphi$.

To see how Markov's Rule is captured by Visser's Scheme, observe that the following formula is one of the consequences of Visser's Scheme,

$$\neg \neg \Box A \rightarrow \Box A.$$

(2.1)

Namely, $\neg \neg \Box A$ is short for $((\Box A \rightarrow \bot) \rightarrow \bot)$, and by Visser's Scheme

$$((\Box A \rightarrow \bot) \rightarrow \bot) \rightarrow ((\Box A \rightarrow \bot)(\Box A, \bot) =$$

$$(\Box A \rightarrow \bot)(\Box A) \vee (\Box A \rightarrow \bot)(\bot) = (\Box A \vee \bot) \leftrightarrow \Box A.$$
Now (2.1) implies that $\text{HA}$ proves the arithmetical realizations of the formula $(\Box \neg \Box A \rightarrow \Box \Box A)$, which is a partial formalization of Markov's Rule. Thus the fact that (2.1) is in the preservativity logic of $\text{HA}$ implies that $\text{HA}$ proves Markov's Rule: $\text{HA} \vdash (\Box \neg \Box A \rightarrow \Box \Box A)$.

We saw that Visser's Scheme describes admissible rules of $\text{HA}$ and considered various consequences of it. In Section 4.6 we will discuss more instances of Visser's Scheme. We will return to the correspondence between preservativity logic and admissible rules in Section 2.3.1.

Summarizing we could say that the preservativity logic presented by Visser (1994) seems a very natural part (if not all) of the preservativity logic of HA. It contains three basic principles, $P1$, $P2$ and Montagna's Principle, which arithmetical validity is trivial. It contains the (preservativity form of the) two characteristic principles of the provability logic of $\text{PA}$, namely $4p$ and $Lp$. And it contains two axioms, the Disjunctive Principle and Visser's Scheme, which are directly related to three well-known properties of $\text{HA}$: the Disjunction Property, Markov's Rule and the propositional admissible rules.

### 2.3.1 Three fragments

Although the preservativity logic of $\text{HA}$ is not known, for three of its fragments there exists a decent axiomatization: for its propositional fragment, for the closed fragment of the provability logic and for that part of the preservativity logic that is connected with the admissible rules of $\text{HA}$.

#### The characterization of the propositional fragment

Recall that $\sigma$ ranges over substitutions which replace propositional variables by arithmetical formulas, and that $\text{IPC}$ denotes intuitionistic propositional logic. It was shown by de Jongh (1982) that

$$\forall \sigma (\text{HA} \vdash \sigma A) \text{iff } \text{IPC} \vdash A.$$  

Note that for propositional formulas, an arithmetical realization $A^*$ is nothing more than a substitution instance $\sigma A$, for some $\sigma$. Therefore, we have

$$\forall \sigma (\text{HA} \vdash \sigma A) \text{iff } \text{IPC} \vdash A.$$  

This means that the propositional fragment of the provability logic of $\text{HA}$ is equivalent to $\text{IPC}$.

#### The characterization of the closed fragment

Visser (1994) described the closed fragment of the provability logic of $\text{HA}$. This is the fragment without propositional variables. He shows that for every formula $\varphi$ in the closed fragment there exists a number $n > 0$ such that

$$\text{HA} \vdash \Box \varphi \leftrightarrow \Box^n \bot.$$
This resembles the situation for PA, where every formula in the closed fragment is a boolean combination of formulas $\Box^n \bot$, $\top$ and $\bot$.

**The characterization of the admissible rules**

The other fragment of the preservativity logic of HA that is axiomatized describes its propositional admissible rules. As mentioned in Section 2.2, if $A \triangleright B$ is in the provability logic of HA, then $A/B$ is an admissible rule of HA. The combination of results by Visser (1994)(1998) and results in part II of this thesis imply that the converse holds too: for all propositional formulas $A, B$ we have that

$$A/B$$ is a propositional admissible rule of HA iff

$$A \triangleright B$$ is in the preservativity logic of HA.

In part II (Chapter 7) of this thesis we give an axiomatization of the propositional admissible rules of HA. In particular, we construct a perspicuous preservativity logic AR such that

$$A/B$$ is a propositional admissible rule of HA iff $AR \vdash A \triangleright B$. \hspace{1cm} (2.2)

This logic is axiomatized by the preservativity principles (Section 2.2) $P1$, $P2$, $Dp$ and all the instances $A \triangleright B$ of $Vp$, where $A$ and $B$ are propositional formulas. In combination with Visser’s (1994) result that states that all these principles belong to the preservativity logic of HA, we arrive at the following axiomatization:

for propositional $A, B$:

$$A \triangleright B$$ is in the preservativity logic of HA iff $AR \vdash A \triangleright B$. \hspace{1cm} (2.3)

This completes our discussion on the three fragments of the preservativity logic of HA for which we have a c.e. axiomatization.

There are two other aspects of (2.2) worth noting. First, it shows that HA *proves* the admissibility of every instance of its propositional admissible rules:

for propositional $A, B$: $A \triangleright B$ is in the preservativity logic of HA iff

$$\forall \sigma (HA \vdash \sigma A \text{ implies } HA \vdash \sigma B)$$ (by definition)

$$A/B$$ is an admissible rule of HA iff (by (2.3) (2.2))

$$\Box A \rightarrow \Box B$$ is in the provability logic of HA iff (by definition)

$$\forall \sigma (HA \vdash \Box \sigma A \rightarrow \Box \sigma B).$$

And second, from (2.2) it follows that

for propositional $A, B$: $A \triangleright B$ is in the preservativity logic of HA iff

$$\Box A \rightarrow \Box B$$ is in the provability logic of HA.

In Section 5.2 of part II we will see that this actually holds for all formulas $(\Box A \rightarrow \Box B)$ of which we know that they are in the preservativity logic of HA. Observe that for example for classical provability principles of the form $(\Box A \rightarrow \Box B)$, like Löb’s Principle $(\Box \Box A \rightarrow A) \rightarrow \Box A$, we already saw that the stronger $A \triangleright B$ holds as well. Of course, the rule does not hold for all arithmetical formulas, as we will see in Section 3.1.
2.4 Interpretability logic

In this section we explain the connection between preservativity logic and interpretability logic. A theory $T$ is $\Pi_1$-conservative over $T'$ if $T'$ proves all the $\Pi_1$-formulas that $T$ proves. From the definition of $\Sigma_1$-preservativity it follows that for classical theories $\Sigma_1$-preservativity is equivalent to $\Pi_1$-conservativity, in the sense that $\varphi \vdash_T \psi$ if and only if $\neg \varphi$ is $\Pi_1$-conservative over $\neg \psi$. For theories that are classical c.e. extensions of PA, this again is equivalent to interpretability (Orey 1961)(Hájek 1971, Hájek 1972). We will not define interpretability here, but remark only that intuitively, '$\varphi$ interprets $\psi$' means that we can define a model for (a translation of) the theory $T$ plus $\psi$ in the theory $T$ plus $\varphi$. Thus for PA we have (in PA) that $\varphi \vdash_{PA} \psi$ if and only if PA plus $\neg \psi$ interprets PA plus $\neg \varphi$.

In a similar manner as for preservativity one can define the interpretability logic of a theory. We denote 'A interprets B' by $A \triangleright_i B$ (in the literature this is denoted by $A \triangleright B$). Interpretability logic has been extensively studied (Shavrukov 1988)(Berarducci 1990)(de Jongh and Veltman 1990)(Zambella 1992) (Visser 1997). The interpretability logic of PA is known to be ILM (a definition follows below). Since for PA the notions of preservativity and interpretability are the same, it seems natural to ask which principles of ILM are inherited by HA. That is, if we reformulate ILM in terms of preservativity by replacing $\neg A \triangleright_i \neg B$ by $A \triangleright B$, which of the principles belong to the preservativity logic of HA?

As we will see, under this translation all axioms of ILM are provable in iPH. Here follow the axioms of ILM. With every axiom we give its preservativity translation. The diamond $\lozenge$ denotes $\square \neg$.

\begin{align*}
L & \quad \square(\square A \rightarrow A) \rightarrow \square A \\
J1 & \quad \square(A \rightarrow B) \rightarrow A \triangleright_i B \\
J2 & \quad A \triangleright_i B \land B \triangleright_i C \rightarrow A \triangleright_i C \\
J3 & \quad A \triangleright_i C \land C \triangleright_i B \rightarrow (A \lor B) \triangleright_i C \\
J4 & \quad A \triangleright_i B \rightarrow (\lozenge A \rightarrow \lozenge B) \\
J5 & \quad \lozenge A \triangleright_i A \\
M & \quad A \triangleright_i B \rightarrow (A \land \square C) \triangleright_i (B \land \square C) \\
4p &
\end{align*}

(The rules of ILM are Modus Ponens and Necessitation.) In Section 3.1 we will see that the translations of $J1$ and $J4$ belong to iPH. Therefore, clearly all translations of ILM belong to iPH.

The converse, i.e. the statement that all translation of axioms of iPH belong to ILM, does not hold. Namely, the translation of $Dp$, which is $C \triangleright_i A \land C \triangleright_i B \rightarrow C \triangleright_i (A \land B)$, is not valid for PA. It is easy to see that the translations of $Vp_n$ and $Lp$ are derivable in PA. Therefore, the only axiom in iPH which does not hold for classical theories is the Disjunctive Principle.
2.5 Intuitionistic provability logic

In this section we explain what is known so far about intuitionistic provability logic and summarize its history.

On the classical side provability logics are well-investigated. On the intuitionistic side we know much less. As stated earlier, it is not known what the provability logic of HA is. The first principles known for this logic were

\[ K \quad \square (A \to B) \to (\square A \to \square B) \]
\[ A \quad \square A \to \square \square A \]
\[ L \quad \square (\square A \to A) \to \square A \quad \text{(Löb's Principle)} \]
\[ Le \quad \square (A \lor B) \to \square (A \lor \square B) \quad \text{(Leivant's Principle)} \]
\[ Ma \quad \square \neg \neg (\square A \to \lor \square B_i) \to \square (\square A \to \lor \square B_i) \quad \text{(Formalized Markov Scheme)} \]

We use the name iH for the logic given by these principles and the rules Modus Ponens and the Necessitation Rule if \( \vdash A \) then \( \vdash \square A \).

In Section 3.1 we show that all these principles and rules are derivable in the preservativity logic iP\( \Pi \) discussed in Section 2.3, and that the latter contains principles not captured by iH. This disproves the conjecture that iH is the provability logic of HA.

The first three principles axiomatize the provability logic \( GL \) of PA, discussed in Section 2.1. Recall (Section 2.2) that Leivant's Principle is related to the Disjunction Property of which the formalized version is not provable in HA.

For the Formalized Markov Scheme as such there is no proof in the literature of its arithmetical validity. Visser (1981) showed that \( \square \neg \neg \square A \to \square \square A \) belongs to the provability logic of HA. From this proof it is not difficult to infer that then also the Formalized Markov Scheme is in the provability logic of HA. This scheme is the partial formalization of Markov's Rule for HA:

for all \( \varphi \in \Pi_2 \): if \( HA \vdash \neg \neg \varphi \), then \( HA \vdash \varphi \).

Clearly any arithmetical realization of formulas of the form \( (\square A \to \lor \square B_i) \) is \( \Pi_2 \).

Arithmetical realizations of formulas \( (\lor \square A_i \to \lor \square B_i) \) are \( \Pi_2 \) too. Note that

\[ \square \neg \neg (\lor \square A_i \to \lor \square B_i) \to \square (\lor \square A_i \to \lor \square B_i) \]

is derivable from the Formalized Markov Scheme. As all formulas not equivalent to a formula of the form \( (\lor \square A_i \to \lor \square B_i) \) have arithmetical realizations that are
not $\Pi_2$, the Formalized Markov Scheme is all we can capture of Markov's Rule in provability logic.

As long as we stay on the classical side (truth) provability logics are very stable; many arithmetical theories have the same provability logic, namely GL (Section 2.1). However, there is in general no stability in going from a classical theory to its intuitionistic counterpart, as can be seen in the comparison of HA and PA. For example, Leivant's Principle $\Box(A \lor B) \rightarrow \Box(A \lor \Box B)$ is not a provability principle of PA. This can be seen easily. If GL would derive the Leivant Principle it would also derive $\Box(\Box \bot \lor \Box \bot)$, as it clearly derives $\Box(\Box \bot \lor \Box \bot)$. But then an application of $L$ shows that it would derive $\Box \bot$. Hence the provability logic of HA is not a part of GL. The converse is not true either. The principle $(p \lor \neg p)$ is a theorem of the provability logic of PA, but not of the corresponding logic of HA. Note that this also shows that there is no monotonicity (converse monotonicity) in provability logics; stronger theories do not necessarily have stronger (weaker) provability logics.

In the context of intuitionistic logic the notion of intuitionistic truth provability logic is less natural, because the intuitionistic notion of truth is much more complex. Therefore, we will in the sequel only discuss provability logic. But let us note in passing that $\Box(A \lor B) \rightarrow \Box A \lor \Box B$ is an example of a principle that is in the truth provability logic of HA but not in the provability logic of HA.

History

The history of intuitionistic provability logic does not reach far back. The first results in this area come from Friedman (1975) and Leivant (1975). As mentioned above, Friedman showed that HA does not prove the formalized version of its Disjunction Property, and Leivant showed that the slightly weaker version $\Box(\varphi \lor \psi) \rightarrow \Box(\varphi \lor \Box \psi)$ is part of the provability logic of HA. Another related result is from Gargov (1984). He has shown that if a c.e. extension of HA has the Disjunction Property then so does its provability logic. Sambin (1976) proved a fixed point theorem for the diagonalizable algebras of intuitionistic theories, which were also studied by Ursini (1979a).

Then there is some work on the algebraic and on the frame characterization of the principles $K$, $4$ and $L$: Ursini (1979b) and Kirov (1984) both show the completeness and the finite model property of iL, and so do Božić and Dösen (1984) for iK, and Wolter and Zakharyaschev (1999b) for iK4.

Of a more arithmetical nature is the paper by Visser (1982). Here he gives some principles of the provability logic of HA, among which the one on the cover of his thesis (Visser 1981): $\Box(\neg \Box A \rightarrow \Box A) \rightarrow \Box \Box A$. All the principles mentioned there are derivable from principles he found later (Visser 1994); they belong to the logic iPH.

The closed fragment of GL and of the provability logic of HA have also been studied. Kirov (1990) shows that the closed fragment of GL is complex in the sense that
any free Heyting algebra with countably many generators can be embedded in the algebra of this fragment. An inspection of the proof shows that this still holds for iPH. As mentioned before, Visser (1994) characterizes the closed fragment of the provability logic of HA by showing that for every formula $\varphi$ in this fragment there exists a number $n > 0$ such that

$$\text{HA} \vdash \Box \varphi \leftrightarrow \Box^n \bot.$$  

Finally, there is the introduction to preservativity logic by Visser (1994)(1998) discussed in Section 2.2 and Chapter 6. Related work has been done by de Jongh and Visser (1996), who studied which c.e. Heyting algebras can be embedded in the Heyting algebras of IPC or HA.

### 2.6 The main roads in intuitionistic modal logic

In this section we introduce intuitionistic modal logic in an informal way, and refer briefly to the different ways in which it has been studied in the literature. The account here is only historical. In Section 3.4 we introduce the modal logics used in this thesis in more detail.

Intuitionistic modal logic is modal logic on an intuitionistic basis. This means that an intuitionistic modal logic is a logic in the language of propositional logic extended with modal operators, that contains IPC. Thus the provability and preservativity logics introduced in the previous sections are modal logics. Intuitionistic polymodal logics have hardly been considered in the literature. Probably this is due to the fact that in the presence of an intuitionistic basis a monomodal logic is almost a bimodal logic; compare the Gödel translation of intuitionistic logic into S4 (Gödel 1933). Most of the logics deal with $\Box$ as well as $\Diamond$, which in general are not interdefinable in an intuitionistic setting. From the point of view of provability logic it is still not clear what a natural interpretation of $\Diamond$ should be. Therefore, in our case we only consider $\Box$ (and $\Rightarrow$).

In the literature on intuitionistic modal logic one often encounters logics of which it is claimed that they are the ‘true’ intuitionistic counterparts of some classical modal logic, for example Löb’s logic. Of course, what one will accept as an intuitionistic counterpart of a given (classical) logic, will depend on the interpretation one has in mind for the modal operators, hence on the properties one wants it to have. In this thesis we always have the provability/preservativity interpretation in mind. A striking difference between this interpretation and most others is that it is in itself of a mathematical nature. Thus verification of the validity of principles can be executed in a formal rigorous way.

Different interpretations of $\Box$ lead to different modal logics. In the literature there have been three prominent perspectives. Prior (1957) first proposed an axiomatization of a modal logic which corresponds to the monadic fragment of intuitionistic predicate logic, by replacing $\Box$, $\Diamond$ and $p_i$ by respectively $\forall x$, $\exists x$ and
Chapter 2. Concepts

Then there are many studies on intuitionistic modal logics whose modal axioms are equivalent to that of a well-known classical system, like K, L, S4 or S5. They contain various possible proof systems (Bierman, Meré and de Paiva 1997) (Simpson 1993), or different possible semantics and completeness results (Fischer-Servi 1977)(Ursini 1979b)(Vakarelov 1981)(Božič and Dösen 1984)(Sotirov 1984) (Simpson 1993)(Wolter and Zakharyaschev 1997)(Wolter and Zakharyaschev 1999a). Fischer-Servi (1977) and Wolter and Zakharyaschev (1999a) also formulate criteria for being the intuitionistic analogue of a classical modal logic. For example, following the definition of Servi, it is not difficult to see that iL is the intuitionistic counterpart of GL. Vakarelov (1981) also shows that above iK there are a continuum of strongly intuitionistic modal logics, i.e. consistent logics that are incompatible with the law of excluded middle. Note that for proper intermediate logics there are none (Rasiowa and Sikorski 1963). Observe that the logic axiomatized by L and Le over iK is only strongly intuitionistic in the weaker sense that it derives □□⊥, see Section 2.5. Modal logics motivated by computer science often turn out to be weaker than iK (Sotirov 1984)(Plotkin and Stirling 1986)(Wijesekera 1990), as do the logics in which the modal operators are viewed as new intuitionistic connectives (Gabbay 1977).

As can be seen from this brief summary of the literature, principles like Le or Vp do not occur, because they neither have a classical counterpart nor do they arise in a natural way from the mentioned interpretations. Thus, looking through the spectacles of provability logic one finds surprising intuitionistic modal logics. Moreover, also on the semantical side certain new possibilities become visible. Besides many other semantics, frame semantics occurs in many of the articles mentioned above. This semantics, defined in Section 3.4.2, consist of a combination of the intuitionistic and the modal frame semantics. That is, frames are sets with two relations: a partial order ≤ (the intuitionistic relation) and a binary relation R (the modal relation). In the presence of only the modal operator □ the canonical frames (Section 3.4.6) satisfy (R; ≤) ⊆ R. Thus it seems harmless to demand this property for frames. However, as we will see in Proposition 4.4.2, some principles can have incompatible frame characterizations with respect to the classes of frames with or without this property. Here again we encounter a deviation from the regular literature on intuitionistic modal logic.

2.7 Arithmetical validity

In this section we prove that the principles and rules given in Section 2.2 and Section 2.5 indeed belong to the preservativity logic of HA. Therefore, the latter belong to the provability logic of HA as well. The main proofs are the one for the Disjunctive Principle and the one for Visser’s Scheme. For the principles Taut, P1, P2 and Montagna’s Principle, these proofs are rather trivial. For the characteristic axioms and rules of the provability logic of PA, namely the principles
2.7. Arithmetical validity

$K$, $4$ and Löb's Principle and the Necessitation Rule, the proofs are analogous to the corresponding proofs for $PA$. Therefore, we do not include them but refer to the literature instead. The proofs for $4p$ and Löb's Preservativity Principle follow easily from the ones for $4$ and Löb's Principle. Finally, we treat the principles that are related to the Disjunction Property and the admissible rules of $HA$ (Section 2.2), the Disjunctive Principle, Visser's Scheme and the Formalized Markov Scheme. For the first two we repeat the proofs by Visser (1994). Then we show that the proof for the last one follows from the fact that Montagna's Principle and Visser's Scheme belong to the preservativity logic of $HA$.

Note that the fact that a principle belongs to the preservativity logic of $HA$ implies that it is arithmetically valid, i.e. the principle holds for $HA$. However, it shows more, namely it shows that $HA$ can also prove this fact.

In this section, we write $\square$ for $\square_H$, and similarly for $\vdash$ and $\vdash$. We will use various properties of $HA$ that hold for $PA$ as well, for example the fact that $HA$ proves $(\theta \rightarrow \square \theta)$ for every $\Sigma_1$-formula $\theta$. We have not included the proofs of these facts, but will refer to the similar proofs for $PA$ in (Hájek and Pudlák 1991) instead. We write $\Gamma \vdash_m \varphi$ for a derivation, in $HA$, of $\varphi$ from $\Gamma$, that uses the finitely many axioms of $\Delta_0 + \text{EXP}$ plus the axioms of $HA$ which Gödel number is smaller than $m$. Similarly for $\square_m$. The reason that we include $\Delta_0 + \text{EXP}$ is that this system is strong enough to allow all coding tricks explained in Section 2.1.

2.7.1. Proposition.

(i) The principles $K$, $4$ and Löb's Principle belong to the provability logic of $HA$ (and hence to its preservativity logic as well).

(ii) Modus Ponens, the Necessitation Rule and the Preservation Rule are rules of the preservativity logic of $HA$ (and hence Modus Ponens and the Necessitation Rule are rules of the provability logic of $HA$ as well).

(iii) The principles Taut, P1, P2 and Montagna's Principle belong to the preservativity logic of $HA$ (and hence Taut belongs to its provability logic as well).

Proof (i) The proofs that $K$, $4$ and Löb's Principle belong to the provability logic of $HA$ are similar to the ones for $PA$, see for example (Smoryński 1985).

(ii) It is trivial that Modus Ponens is a rule of the preservativity logic of $HA$, because it is a rule of the logic of $HA$. The proof that $HA$ satisfies the Necessitation Rule, if $HA \vdash \varphi$ then $HA \vdash \square \varphi$, is similar to the one for $PA$, see (Smoryński 1985). The fact that $HA$ satisfies the Preservation Rule, if $HA \vdash (\varphi \rightarrow \psi)$ then $HA \vdash \varphi \triangleright \psi$, follows almost immediately. Suppose $HA \vdash (\varphi \rightarrow \psi)$. Hence by Necessitation Rule we have $HA \vdash \square (\varphi \rightarrow \psi)$. It is easy to see that this implies $HA \vdash \varphi \triangleright \psi$. In the next chapter, Section 3.1, we will see that there is an equivalent formulation of preservativity logic for which the Preservation Rule is replaced by the Necessitation Rule.
(iii) The statement that $Taut$ belongs to the preservativity logic of HA is trivial, since the logic of HA is intuitionistic predicate logic, which contains IPC. The proofs for $P1$ and $P2$ are left to the reader. For Montagna’s Principle, consider formulas $A, B, C$ and an arithmetical translation $*$. We have to show that

$$HA \vdash A^* \rightarrow B^* \rightarrow (\Box C^* \rightarrow A^*) \rightarrow (\Box C^* \rightarrow B^*).$$

Recall that the arithmetical realization of $\Box C$ is a $\Sigma_1$-formula (Section 2.1). Therefore, it suffices to show that for all arithmetical formulas $\varphi, \psi$ we have

$$(2.4) \quad \text{for all } \Sigma_1\text{-formulas } \theta: HA \vdash \varphi \rightarrow (\theta \rightarrow \varphi) \rightarrow (\theta \rightarrow \psi).$$

In fact, HA even proves: for all $\Sigma_1$-formulas $\varphi$, $\varphi \rightarrow \psi$ implies $(\theta \rightarrow \varphi) \rightarrow (\theta \rightarrow \psi)$. As we do not need this stronger statement, we prove the weaker $(2.4)$ instead. We use that if a formula is $\Sigma_1$, then HA proves this fact, and that HA proves that $\Sigma_1$-formulas are closed under conjunction. These properties of HA are proved in a similar way as for PA, see (Hájek and Pudlák 1991).

The proof of $(2.4)$ runs as follows. Let $\theta$ be a $\Sigma_1$-formula. Reason in HA. Suppose $\varphi \rightarrow \psi$. We have to prove that for all $\Sigma_1$-formulas $\theta'$, if $\vdash (\theta' \rightarrow (\theta \rightarrow \varphi))$ holds, then $\vdash (\theta' \rightarrow (\theta \rightarrow \psi))$ holds as well. Therefore, suppose $\vdash (\theta' \rightarrow (\theta \rightarrow \varphi))$, for some $\Sigma_1$-formula $\theta'$. Note that $(\theta' \rightarrow (\theta \rightarrow \varphi))$ is equivalent to $(\theta' \land \theta \rightarrow \varphi)$. Thus by Necessitation Rule (ii) and the axiom $K (i)$, we also have that $\vdash (\theta' \rightarrow (\theta \rightarrow \varphi))$ is equivalent to $\vdash (\theta' \land \theta \rightarrow \varphi)$. The conjunction of two $\Sigma_1$-formulas is a $\Sigma_1$-formula, and whence $(\theta' \land \theta \rightarrow \varphi)$ is a $\Sigma_1$-formula. Therefore, by $\varphi \rightarrow \psi$, $\vdash (\theta' \land \theta \rightarrow \varphi)$ implies $\vdash (\theta' \land \theta \rightarrow \psi)$. The latter is again equivalent to $\vdash (\theta' \rightarrow (\theta \rightarrow \psi))$, which completes the proof. \hfill \Box

2.7.2. Proposition. The principle $4p$ and Löb’s Preservativity Principle belong to the preservativity logic of HA.

Proof In Section 4.3 we show that Löb’s Preservativity Principle derives the principle $4p$. Therefore, it suffices to show that Löb’s Preservativity Principle is a principle of the preservativity logic of HA. We use the well-known fact that, like PA, HA proves $\Sigma_1$-completeness i.e. HA proves that for every $\Sigma_1$-formula $\theta$ we have $HA \vdash (\theta \rightarrow \Box \theta)$. A proof for PA, which is analogous to the one for HA, can be found in (Hájek and Pudlák 1991).

Reason in HA. If for some $\theta \in \Sigma_1$ we have $\vdash (\theta \rightarrow (\Box \varphi \rightarrow \varphi))$, then we also have $\vdash (\Box \theta \rightarrow (\Box \varphi \rightarrow \varphi))$ by the Necessitation Rule and the axiom $K$ (Proposition 2.7.1). Since $(\theta \rightarrow \Box \theta)$ by $\Sigma_1$-completeness, also $\vdash (\theta \rightarrow (\Box \varphi \rightarrow \varphi))$. Applying Löb’s Principle (Proposition 2.7.1) gives $\vdash (\theta \rightarrow \Box \varphi)$. Thus by assumption also $\vdash (\theta \rightarrow \varphi)$. \hfill \Box

The proofs (Visser 1994) that the Disjunctive Principle and Vissers’s Scheme belong to the preservativity logic of HA are related but not similar. This difference
2.7. Arithmetical validity

is not surprising, as in contrast to the Disjunctive Principle, Visser's Scheme is classically valid, i.e. it belongs to the preservativity logic of PA (see Section 2.3). Before giving those proofs, we will briefly sketch the ideas behind them. They both use a translation on formulas by D. de Jongh. Translations, like for example realizability or the Friedman translation, are a much used tool in meta proofs for constructive theories. In such proofs one often shows that if a formula is derivable then the translation of that formula is also derivable. In our case, we proceed in a similar way. We construct some kind of \( \Sigma_1 \)-approximations to the formulas involved, and use the de Jongh translation to show that if the original formula is derivable then these \( \Sigma_1 \)-approximations have the desired properties.

In the case of the Disjunctive Principle we have to show, in HA, that if \( \varphi \vdash \chi \) and \( \psi \vdash \chi \) hold, then also \( (\varphi \lor \psi) \vdash \chi \). Thus we have to prove, in HA, that for all \( \Sigma_1 \)-formulas \( \theta \) with \( \vdash (\theta \rightarrow (\varphi \lor \psi)) \), we have \( \vdash (\theta \rightarrow \chi) \). It suffices to show that for every \( \Sigma_1 \)-formula \( \theta \) with \( \vdash (\theta \rightarrow (\varphi \lor \psi)) \), we can find \( \Sigma_1 \)-formulas \( \theta_i \), the disjunction of which is implied by \( \theta \), and such that \( \vdash (\theta_i \rightarrow \varphi) \) and \( \vdash (\theta_2 \rightarrow \psi) \). Namely, in that case \( \varphi \vdash \chi \) implies \( \vdash (\theta_1 \rightarrow \chi) \) and similarly for \( \psi \).

We will see that for some \( m \), we can take the formulas \( \Box_m \varphi \) and \( \Box_m \psi \) for \( \theta_i \):

\[ (i) \text{ for } \Sigma_1\text{-formula } \theta: \vdash (\theta \rightarrow (\varphi \lor \psi)) \text{ implies } \vdash (\theta \rightarrow \Box_m \varphi \lor \Box_m \psi) \]

\[ (ii) \vdash (\Box_m \varphi \rightarrow \varphi) \text{ and } \vdash (\Box_m \psi \rightarrow \psi). \]

Only one of these statements has to do with the constructive properties of HA. Namely, \( (ii) \) holds for PA as well, while \( (i) \) does not. For the latter, this is easy to see. Consider the case \( \psi = \neg \varphi \) and \( \theta = \top \). Then \( (i) \) would show that for all \( \varphi \), PA derives \( \Box_m \varphi \lor \Box_m \neg \varphi \), a fact which is not even true. However, we will see that in the context of HA both properties hold and this will complete the proof.

As mentioned before, Visser's Scheme is classically valid, and we will see that PA occurs in the proof that Visser's Scheme belongs to the preservativity logic of HA. Namely, we use the well-known fact that PA is \( \Pi_2 \)-conservative over HA, and that HA proves this fact (Friedman 1977). To explain the idea of this proof, consider the following instance of Visser's Scheme:

\[ ((\varphi_1 \rightarrow \psi) \rightarrow \varphi_2) \vdash ((\varphi_1 \rightarrow \psi)(\varphi_1) \lor (\varphi_1 \rightarrow \psi)(\varphi_2)). \]

We have to show, in HA, that for all \( \Sigma_1 \)-formulas \( \theta \) it holds that

\[ \vdash \theta \rightarrow ((\varphi_1 \rightarrow \psi) \rightarrow \varphi_2)) \text{ implies } \vdash \theta \rightarrow (\varphi_1 \rightarrow \psi)(\varphi_1) \lor (\varphi_1 \rightarrow \psi)(\varphi_2). \]

We consider only the case that \( \theta = \top \). Therefore, suppose

\[ \vdash ((\varphi_1 \rightarrow \psi) \rightarrow \varphi_2)). \quad (2.5) \]

We have to show that

\[ \vdash (\varphi_1 \rightarrow \psi)(\varphi_1) \lor (\varphi_1 \rightarrow \psi)(\varphi_2). \quad (2.6) \]
Note that in the case that the formulas $\varphi_i$ are not of the form $\Box \varphi'$, the fact that (2.5) implies (2.6), expresses a well-known admissible rule of HA. Therefore, the following proof sketch shows that HA recognizes this admissible rule (compare the part of Section 2.2 on Visser’s Scheme).

Since HA is part of PA, the latter derives (2.5) too. By classical reasoning it follows that

$$\text{PA} \vdash (\varphi_1 \lor \varphi_2).$$

If $(\varphi_1 \lor \varphi_2)$ is $\Pi_2$ then by the mentioned $\Pi_2$-conservativity of PA over HA, we can conclude that HA derives this formula. This already explains the instance of Visser’s Scheme for which the formulas $\varphi_i$ are of the form $\varphi = \Box \varphi'$, and hence $\Sigma_1$.

However, the formula $(\varphi_1 \lor \varphi_2)$ is not $\Pi_2$ in general. Therefore, we have to find some kind of $\Sigma_1$-approximation of $\varphi_i$, which means a $\Sigma_1$-formula $\varphi'_i$ such that $((\varphi_i \rightarrow \varphi')$ and $(\varphi'_i \rightarrow (\varphi_1 \rightarrow \psi)(\varphi_i))$. Namely, in that case (2.5) implies that $\vdash ((\varphi'_i \rightarrow \psi) \rightarrow \varphi'_2)$. And the same reasoning as above shows that PA derives $(\varphi'_1 \lor \varphi'_2)$. Since this is a $\Pi_2$-formula, by $\Pi_2$-conservativity we can conclude that HA $\vdash (\varphi'_1 \lor \varphi'_2)$. Using the other property of $\varphi'_i$ we arrive at the desired conclusion (2.6).

As we will see, these formulas $\varphi'_i$ actually do not have the property $(\varphi_i \rightarrow \varphi')$. However, using the de Jongh translation we can show that (2.5) implies that for some $\psi'$, $\vdash ((\varphi'_i \rightarrow \psi') \rightarrow \varphi'_2)$ holds. Then we reason as before and get (2.6) as well.

The properties of $\Sigma_1$-formulas in the previous discussion already hints at the special treatment of formulas of the form $\Box C$, which arithmetical translations are $\Sigma_1$, in Visser’s Scheme.

Before giving the formal proofs of the two principles discussed above, we need some definitions and lemmas. The translation on arithmetical formulas by D. de Jongh, is given by the following inductive definition.

\[ [\chi]_m(\varphi) \equiv_{\text{def}} \varphi, \text{ for atomic } \varphi \]

\[ [\chi]_m(\cdot) \text{ commutes with } \land, \lor, \exists \]

\[ [\chi]_m(\varphi \rightarrow \psi) \equiv_{\text{def}} ([\chi]_m(\varphi) \rightarrow [\chi]_m(\psi)) \land [\chi]_m(\varphi \rightarrow \psi) \]

\[ [\chi]_m(\forall x \varphi x) \equiv_{\text{def}} \forall x ([\chi]_m(\varphi x) \land [\chi]_m(\varphi x \rightarrow \forall x \varphi x)). \]

We write $[\chi]_m(\Gamma)$ for \{ $[\chi]_m(\psi) \mid \psi \in \Gamma$ \}. Define

\[ (\chi)_m(\varphi) \equiv_{\text{def}} \varphi, \text{ for atomic } \varphi \]

\[ (\chi)_m(\cdot) \text{ commutes with } \land, \lor, \exists \]

\[ (\chi)_m(\varphi \rightarrow \psi) \equiv_{\text{def}} \Box_m(\chi \rightarrow (\varphi \rightarrow \psi)) \]

\[ (\chi)_m(\forall x \varphi x) \equiv_{\text{def}} \Box_m(\chi \rightarrow \forall x \varphi x) \]
2.7.3. Lemma. We have, verifiably in HA, that

\[ [\chi]_m(\varphi) \rightarrow (\chi)_m(\varphi). \]
\[ [\chi]_m(\varphi) \rightarrow \Box_m(\chi \rightarrow \varphi). \]

\((\chi)_m(\varphi) \in \Sigma_1.\)

for all \(\theta \in \Sigma_1;\) \([\chi]_m(\theta) \leftrightarrow \theta \leftrightarrow (\chi)_m(\theta).\)

Proof Using induction on \(\varphi,\) the proofs of the first and the third statement are straightforward. For the last equation, use the fact that HA derives the formula \(\forall x \leq y \Box_m(\varphi x) \rightarrow \Box_m \forall x \leq y(\varphi x)\) (a proof for PA can be found in (Hájek and Pudlák 1991)). This implies that HA \(\vdash [\chi]_m(\forall x \leq y \varphi x) \leftrightarrow \forall x \leq y[\chi]_m(\varphi x),\)

and the rest of the statement follows easily. The proof of the second statement follows from the fact that, verifiably in HA, \(\Delta_0 + \text{EXP}\) proves \(\Sigma_1\)-completeness: for \(\Sigma_1\)-formula \(\theta\) it holds that \((\theta \rightarrow \Box \theta).\) An analogous proof for PA can be found in (Hájek and Pudlák 1991). Once this is known, the rest of the proof is easy. \(\square\)

2.7.4. Lemma. For all formulas \(A, B\) in preservativity logic, for all arithmetical realizations \(*\), and for all \(m,\) we have

\(\text{HA} \vdash (A^*)_m(B^*) \rightarrow (A^*)(B^*).\)

Proof First note that for all natural numbers \(m,\) HA proves \(\Box_m \varphi \rightarrow \varphi.\) The proof is completely similar to the one for PA (Hájek and Pudlák 1991). Recall the definition of \((A)(B_1, \ldots, B_n)\) for the case \(n = 1:\)

\[(A)(\bot) \equiv_{\text{def}} \bot\]
\[(A)(B \land B') \equiv_{\text{def}} (A)(B) \land (A)(B')\]
\[(A)(\Box B) \equiv_{\text{def}} \Box B\]
\[(A)(B) \equiv_{\text{def}} (A \rightarrow B),\text{ for } B\text{ not of the form }\bot, (C \land C') \text{ or } \Box C.\]

For all these cases we have to prove that HA derives \((A^*)_m(B^*) \rightarrow (A^*)(B^*).\)

Reason in HA. From the definition of \((\chi)_m(\varphi)\) and Lemma 2.7.3 it follows that we have,

\((\chi)_m(\bot) \leftrightarrow \bot\]
\[(\chi)_m(\varphi) \leftrightarrow \varphi,\text{ if } \varphi\text{ is a }\Sigma_1\text{-formula.}\]

We show that \((A^*)_m(B^*) \rightarrow (A)(B)^*\) holds with induction to \(B.\) In the case that \(B = \bot\) it is easy to see that \((A^*)_m(B^*) \rightarrow (A)(B)^*.\)

If \(B = \Box C,\) then \(B^*\) is a \(\Sigma_1\)-formula. Hence it holds that \((A^*)_m(B^*) \leftrightarrow \Box C^*,\) and \(\Box C^* = (A^*)(B^*).\) If \(B = C \supset D,\) then \(B^*\) is of the form \(\forall x \varphi x,\) because \(C^* \supset D^*\) says 'for all \(x,\) if \(x\) is the code of a \(\Sigma_1\)-formula \(\theta\) and \(\Box (\theta \rightarrow C^*)\) holds, then
\( \Box(\theta \rightarrow D^*) \) holds as well. Thus \((A^*)_m(B^*) = \Box_m(A^* \rightarrow B^*) \). By the observation above this implies that we have \((A^* \rightarrow B^*) \), which is \((A^*)(B^*) \).

If \( B = (C \land D) \), then \((A^*)_m(B^*) = (A^*)_m(C^*) \land (A^*)_m(D^*) \). By the induction hypothesis, \((A^*)_m(C^*) \land (A^*)_m(D^*) \) implies \((A^*)(C^*) \land (A^*)(D^*) \). By definition, \((A^*)(C^*) \land (A^*)(D^*) = (A^*)(C^* \land D^*) = (A^*)(B^*) \).

If \( B = (C \lor D) \), then \((A^*)_m(B^*) = (A^*)_m(C^*) \lor (A^*)_m(D^*) \). By the induction hypothesis, \((A^*)_m(C^*) \lor (A^*)_m(D^*) \) implies \((A^*)(C^*) \lor (A^*)(D^*) \). It is easy to see that \((A^*)(C^*) \lor (A^*)(D^*) \) implies \((A^*)(C^* \lor D^*) \) (Lemma 3.2.1 (i)).

If \( B = (C \rightarrow D) \), then \((A^*)_m(B^*) = \Box_m(A^* \rightarrow B^*) \). By the observation above this gives \((A^* \rightarrow B^*) \), which is \((A^*)(B^*) \). 

\( \Box \)

2.7.5. Lemma. (Visser 1994) Let \( \varphi = \bigwedge_{i=1}^n (\varphi_i \rightarrow \psi_i) \). We have, verifiably in HA,

\[
\llbracket \chi \rrbracket_m(\varphi) \leftrightarrow (\llbracket \chi \rrbracket_m(\varphi_i) \rightarrow \llbracket \chi \rrbracket_m(\psi_i)) \land \Box_m(\chi \rightarrow \varphi)
\]

\( \Gamma \vdash_{m} \psi \) implies \( \llbracket \chi \rrbracket_m(\Gamma) \vdash \llbracket \chi \rrbracket_m(\psi) \).

**Proof** The proof of the first equation is left to the reader. For the second statement we use induction to the length of the derivation \( \Gamma \vdash_{HA,m} \varphi \). We treat the two difficult cases:

Case 1. \( \Gamma \) is empty and \( \varphi = \psi_0 \land \forall x(\psi x \rightarrow \psi(x + 1)) \) \( \rightarrow \forall x \psi x \), i.e. \( \varphi \) is an induction axiom. Since \( \vdash \Box m \varphi \), also \( \Box_m(\chi \rightarrow \varphi) \). It remains to show that we have \( \vdash \llbracket \chi \rrbracket_m(\psi_0 \land \forall x(\psi x \rightarrow \psi(x + 1))) \rightarrow \llbracket \chi \rrbracket_m(\forall x \psi x) \), which is equivalent to

\[
\llbracket \chi \rrbracket_m(\psi_0) \land \forall x(\llbracket \chi \rrbracket_m(\psi x \rightarrow \llbracket \chi \rrbracket_m(\psi(x + 1))) \land \\
\Box_m(\chi \rightarrow \forall x(\psi x \rightarrow \psi(x + 1))) \rightarrow \forall x \llbracket \chi \rrbracket_m(\psi x) \land \Box_m(\chi \rightarrow \forall x \psi x).
\]

As we observed in Lemma 2.7.3, \( \llbracket \chi \rrbracket_m(\psi_0) \) implies \( \Box_m(\chi \rightarrow \psi_0) \). Hence from \( \Box_m(\chi \rightarrow \varphi) \) it follows that \( \Box_m(\chi \rightarrow \forall x(\psi x \rightarrow \psi(x + 1))) \) implies \( \Box_m(\chi \rightarrow \forall x \psi x) \).

By induction we have

\[
\llbracket \chi \rrbracket_m(\psi_0) \land \forall x(\llbracket \chi \rrbracket_m(\psi x \rightarrow \llbracket \chi \rrbracket_m(\psi(x + 1))) \rightarrow \\
\forall x \llbracket \chi \rrbracket_m(\psi x) \land \Box_m(\chi \rightarrow \forall x \psi x).
\]

And this concludes Case 1.

Case 2. Suppose \( \varphi = (\psi \rightarrow \psi') \) and the last step in the proof is \( \Gamma, \psi \vdash_{m} \psi' \) implies \( \Gamma \vdash_{m} (\psi \rightarrow \psi') \). By the induction hypothesis, \( \Gamma, \psi \vdash_{m} \psi' \) implies \( \llbracket \chi \rrbracket_m(\Gamma), \llbracket \chi \rrbracket_m(\psi) \vdash_{HA} \llbracket \chi \rrbracket_m(\psi') \). And thus

\[
\llbracket \chi \rrbracket_m(\Gamma) \vdash \llbracket \chi \rrbracket_m(\psi) \rightarrow \llbracket \chi \rrbracket_m(\psi').
\]

Therefore, it remains to show that

\[
\llbracket \chi \rrbracket_m(\Gamma) \vdash \Box_m(\chi \rightarrow (\psi \rightarrow \psi')).
\]

Clearly, we have \( \theta, \psi \vdash_{m} \psi' \) for some conjunction \( \theta \) of elements of a finite subset of \( \Gamma \). Thus we have \( \llbracket \chi \rrbracket_m(\Gamma) \vdash \Box_m(\chi \rightarrow \theta) \) and \( \Box_m(\theta \rightarrow (\psi \rightarrow \psi')) \). And this leads to the desired conclusion. 

\( \Box \)
2.7.6. Theorem. (Visser 1994) The Disjunctive Principle belongs to the preservativity logic of HA.

Proof It is a well-known fact that, like PA, HA proves reflection for its finite fragments, i.e. for every natural number n, HA proves (□_n φ → φ). Moreover, HA can prove this fact, that is, HA proves that for every x, ⊩ (□_x φ → φ). A proof of this fact for PA, which is similar to the one for HA, can be found in (Hájek and Pudlák 1991). The proof that the Disjunctive Principle belongs to the preservativity logic of HA now runs as follows.

Reason in HA. Suppose that φ ⊩ χ and ψ ⊩ χ hold. We have to show that (φ ∨ ψ) ⊩ χ holds, i.e. that for all Σ₁-formulas θ, ⊩ (θ → (φ ∨ ψ)) implies ⊩ (θ → χ). Therefore, consider a Σ₁-formula θ and suppose ⊩ (θ → (φ ∨ ψ)). Thus θ ⊩ (φ ∨ ψ), for some m. By Lemma 2.7.5 we have

\[ [T]_m(θ) ⊩ [T]_m(φ) \lor [T]_m(ψ). \]

By Lemma 2.7.5 this implies

\[ θ ⊩ □_m φ \lor □_m ψ. \]  \( (2.7) \)

Note that □_m φ and □_m ψ are Σ₁-formulas. As observed above, we have that

\[ ⊩ (□_n φ → φ) \text{ and } ⊩ (□_n ψ → ψ). \]

Therefore, from φ ⊩ χ and ψ ⊩ χ, we conclude

\[ ⊩ (□_m φ → χ) ∧ (□_m ψ → χ). \]

Together with (2.7) this gives

\[ ⊩ (θ → χ). \]

This completes our proof. □

2.7.7. Theorem. (Visser 1994) Visser’s Scheme belongs to the preservativity logic of HA.

Proof We have to show that for all arithmetical formulae φᵢ, ψᵢ, for all n, if

\[ χ = \bigwedge_{i=1}^{n}(φᵢ → ψᵢ), \]

then we have

\[ HA ⊩ (χ → φ_{n+1} ∨ φ_{n+2}) ⊩ (\bigvee_{i=1}^{n}(χ)_m(φᵢ)). \]  \( (2.8) \)

Clearly, this implies that for all formulas Aᵢ, Bᵢ in the language of preservativity and for all arithmetical translations * , if A = \bigwedge_{i=1}^{n}(Aᵢ → Bᵢ), then

\[ HA ⊩ (A^* → A^*_{n+1} ∨ A^*_{n+2}) ⊩_{HA} (\bigvee_{i=1}^{n}(A^*)_m(A^*_ᵢ)). \]
By Lemma 2.7.4, \( \text{HA} \) derives \((A^*)_m(A^*_i) \rightarrow (A^*)(A^*_i)\). Thus by Preservation Rule (Proposition 2.7.1), \( \text{HA} \) derives \((A^*)_m(A^*_i) \vdash (A^*)(A^*_i)\). Applying the principle PI (Proposition 2.7.1) now gives

\[
\text{HA} \vdash (A^* \rightarrow A^*_{n+1} \lor A^*_{n+2}) \vdash (\bigvee_{i=1}^{n+2} (A^*_i)(A^*_i)).
\]

Using the fact that \((\bigvee_{i=1}^{n+2} (A^*_i)(A^*_i)) = (A^*)(A^*_1, \ldots, A^*_{n+2})\), this implies Visser’s Scheme:

\[
\text{HA} \vdash (A^* \rightarrow A^*_{n+1} \lor A^*_{n+2}) \vdash (A^*)(A^*_1, \ldots, A^*_{n+2}).
\]

Therefore, to show that Visser’s Scheme belongs to the preservativity logic of \( \text{HA} \) it suffices to show that (2.8) holds, i.e. that \( \text{HA} \) derives that for all \( \theta \in \Sigma_1 \), \( \vdash (\theta \rightarrow (X \rightarrow \varphi_{n+1} \lor \varphi_{n+2})) \) implies \( \vdash (\theta \rightarrow (X)(\varphi_1, \ldots, \varphi_{n+2})) \).

Reason in \( \text{HA} \). Let \( \theta \in \Sigma_1 \) and assume \( \vdash (\theta \rightarrow (X \rightarrow \varphi_{n+1} \lor \varphi_{n+2})) \). Hence for some \( m \), we have \( \vdash m (X \rightarrow \varphi_{n+1} \lor \varphi_{n+2}) \). From Lemma 2.7.5 it follows that \( [\chi]_m(\theta) = [\chi]_m(\varphi_{n+1} \lor \varphi_{n+2}) \). Hence by the same lemma:

\[
\theta \vdash \bigwedge_{i=1}^{n} ([\chi]_m(\varphi_i) \rightarrow [\chi]_m(\psi_i)) \land \Box_n (X \rightarrow X) \rightarrow [\chi]_m(\varphi_{n+1} \lor \varphi_{n+2}).
\]

Thus clearly,

\[
\theta \vdash \bigwedge_{i=1}^{n} ([\chi]_m(\varphi_i) \rightarrow [\chi]_m(\psi_i)) \rightarrow [\chi]_m(\varphi_{n+1} \lor \varphi_{n+2}).
\]

By Lemma 2.7.3 and elementary reasoning this implies that

\[
\theta \vdash \bigwedge_{i=1}^{n} ((\chi)_m(\varphi_i) \rightarrow [\chi]_m(\psi_i)) \rightarrow (\chi)_m(\varphi_{n+1} \lor \chi)_m(\varphi_{n+2}).
\]

Hence

\[
\text{PA} \vdash \theta \rightarrow \bigwedge_{i=1}^{n} ((\chi)_m(\varphi_i) \rightarrow [\chi]_m(\psi_i)) \rightarrow (\chi)_m(\varphi_{n+1} \lor (\chi)_m(\varphi_{n+2})).
\]

Using classical logic we can conclude that \( \text{PA} \vdash \theta \rightarrow \bigvee_{i=1}^{n+2} (\chi)_m(\varphi_i) \). By Lemma 2.7.3, \( \theta \rightarrow \bigvee_{i=1}^{n+2} (\chi)_m(\varphi_i) \) is a \( \Pi_2 \)-formula. By the \( \Pi_2 \)-conservativity of \( \text{PA} \) over \( \text{HA} \) mentioned above, we have

\[
\vdash \theta \rightarrow \bigvee_{i=1}^{n+2} (\chi)_m(\varphi_i).
\]

This completes the proof of (2.8). \( \square \)
2.8. **Overview of part I**

A **Corollary.** The Formalized Markov Scheme belongs to the provability logic of HA (and hence to its preservativity logic).

Proof In Chapter 3 (Section 3.3) we show that the Formalized Markov Scheme is derivable from Visser’s Scheme and Montagna’s Principle, using the rules Modus Ponens and Necessitation.


2.8 **Overview of part I**

Part I of this thesis is a modal study of the principles of the preservativity logic of HA known so far. In particular, we prove the frame completeness of the conjectured preservativity logic iP\(H\) of HA, the main result of this part of the thesis (Chapter 5). As explained in Section 2.1 such a characterization is often the first step for finding embeddings of a provability logic in the corresponding arithmetical theory, i.e. for showing that a system is the provability logic of some theory. We also show in Chapter 5 that the system iP\(H\) contains principles of the provability logic of HA that are not captured by i\(H\). This disproves the conjecture that i\(H\) is the provability logic of HA.

The proofs in Chapter 5 use the results of Chapter 4, where we study the principles separately. Here we also show that besides the principles \(V_p_n\) all principles are independent, as expected. Moreover, there we will see that Visser’s Scheme is infinite in an essential way: it is not equivalent to a finite number of Visser’s Principles.

As mentioned in the introduction, the characterization of the principles requires many technical tools from modal logic. Moreover, these logics deviate a lot from the logics that are regularly studied in intuitionistic modal logic. Whence some surprising properties and problems become visible, and many proofs are quite different from the ones for the modal logics one usually encounters. Therefore, also from the modal point of view these logics are interesting.

Chapter 7 of part II of this thesis could also be seen in the light of provability logic. This was explained in Section 2.3.1 where we discussed three particular fragments of the preservativity logic of HA.

In Chapter 3 we introduce the tools used in the following chapters of part I. Section 3.4 contains preliminaries. In Section 3.3 we show how the principles of the provability logic of HA, i.e. of the logic i\(H\), are captured by its conjectured preservativity logic iP\(H\).