Provability interpretation
of propositional and modal logics

Amirhossein Akbar Tabatabai

Provability Interpretation of Propositional and Modal Logics

Amirhossein Akbar Tabatabai [*]
Institute of Mathematics
Academy of Sciences of the Czech Republic
tabatabai@math.cas.cz

Abstract

In 1933 [6], Gödel introduced a provability interpretation of the propositional intuitionistic logic to establish a formalization for the BHK interpretation. He used the modal system, **S4**, as a formalization of the intuitive concept of provability and then translated **IPC** to **S4**. His work suggested the problem to find a concrete provability interpretation of the modal logic **S4**. In this paper, we will try to answer this problem. In fact, we will generalize Solovay's provability interpretation of the modal logic **GL** to capture other modal logics like **K4**, **KD4** and **S4**. Then we will use these results to find a formalization for the BHK interpretation and we will show that with different interpretations of the BHK interpretation, we can capture some of the propositional logics like Intuitionistic logic, minimal logic and Visser-Ruitenburg's basic logic.

Moreover, we will show that there is no provability interpretation for any extension of **KD45** and also there is no BHK interpretation for the classical propositional logic.

1 Introduction

BHK interpretation

In the intuitionistic tradition, mathematics is considered as a theory of mental constructions and hence, truth naturally means the existence of a proof. Thus, provability is the core stone of the whole intuitionistic paradigm. With

this fact in mind, like any other logic, the intuitionistic logic would be a calculus to describe the behavior of truth, which in this case, is the concept of provability. In other words, intuitionistic logic is a meta-theory of the concept of provability. Now let us explain the role of connectives in this logic. Again like any other logic, a connective is an operation on the truth content of its inputs, which in the case of intuitionistic logic means the operations on the proofs. Now, if we want an intuitive semantics for intuitionistic logic, we have to find out what the meaning of a connective is. The answer to this question is the well-known BHK interpretation. Its propositional part is the following:

- a proof for $A \wedge B$ is a pair of a proof for A and a proof for B.
- a proof for $A \vee B$ is a proof for A or a proof for B.
- a proof for $A \rightarrow B$ is a construction which transforms any proof of A to a proof for B.
- a proof for $\neg A$ is a construction which transforms any proof of A to a proof for $\bot$.
- $\bot$ does not have any proof.

Clearly, what we proposed as the BHK interpretation is just an informal interpretation and we need to find its exact formalization, if we want to use it as a mathematical tool. For instance, if we want to establish an argument which shows that Heyting's formalization of **IPC** is an adequate formalization for intuitionistic viewpoint, we have to prove the soundness and completeness of **IPC** with respect to the BHK interpretation and it obviously needs its exact formalized version. Now, to formalize the interpretation, we firstly need a formalization of the concept of proof and based on the extensive works in proof theory, which have been done so far, it seems quite possible to find an appropriate formalization for the proof and hence for the BHK interpretation. But unfortunately, despite all the attempts which have been made, the BHK interpretation has not been formalized so far (for an extensive history of the problem see [1]). Why does this natural and simple interpretation resist to become formalized? To find an answer to the question, let us investigate one of the key properties of the interpretation. Think about a proposition like $A \rightarrow B$. Its proof is a construction which transforms any proof of A to a proof of B. It is clear that this construction would be a meta-proof and not just a proof, because it talks about proofs and therefore it should belong to the meta-language of A and B. In other words, we could claim that the act of introducing an implication increases the layer of the meta-language which we argue in. Therefore, in BHK interpretation all levels of our meta-languages involve and this is the reason why this interpretation is so complex to be formalized. Now, since we need

to formalize the meaning of a proof, we have to extend our task to find a meaning of a proof in any level of the meta-languages.

There are two different approaches to implement this idea. In the first approach, we could be faithful to the intuitionistic paradigm and find an intuitionistically valid interpretation of the proofs. However, in the second approach we could change our viewpoint and construct a bridge to find an appropriate classical interpretation of the concept of a proof to formalize the BHK interpretation. The first approach is Heyting's approach and the second one is Kolmogorov's. At the first glance, the first approach seems very natural to try but there is a huge problem there; a conceptual vicious circle which forces us to understand the semantics of the paradigm, the BHK interpretation, in terms of itself and it makes the whole process very complicated. We want to emphasize that this vicious circle does not mean that the first approach is philosophically invalid, but it just shows how complex it could be. (Think about classical logic and its semantics which is based on the classical meta-theory and this is an obvious vicious circle. But these kinds of vicious circles are the inherent properties of any paradigm in the philosophy of mathematics and they should not be avoided.) In this paper we follow the second approach and interpret all proofs as the classical proofs in different layers of meta-languages. But this is not an easy task to do and in the forthcoming part of the introduction we will investigate the problems in this approach.

The last thing we wanted to mention here is that what we are going to formalize, is actually an implicit version of the BHK interpretation, instead of the original one. In the original interpretation we interpret all the connectives as operations on explicitly mentioned proofs. But we could somehow eliminate the *proofs* from the interpretation and just talk about the *provability* of a sentence. For instance, the disjunction case in the original BHK interpretation transforms to the following one: $A \vee B$ is provable if A is provable or B is provable. The problem here, is the case of implication which is not reducible to a simpler one. In order to solve this problem, we need a primitive connective to formalize the concept of provability. A role which would be played by the connective "box" in modal logics and this is one of the most important contributions to the problem, which was made by Kurt Gödel. Now, Gödel's contribution.

Gödel's Translation

In 1933 [6], Gödel introduced a provability interpretation of **IPC** that can be seen as an implicit version of the well-known BHK interpretation of the intuitionistic logic. By this interpretation he could justify the fact that Heyting's formalization of **IPC** is sound and complete for its intended semantics which is the BHK interpretation. Let us review some steps of his work.

1. Giving a proof interpretation: Before Giving any provability interpretation of **IPC**, we should explain our intention of the concept of *provability* and the properties that we want to have. As you expect, Gödel began his work exactly from this point. He used the language of modal logics, in which the symbol "□" is interpreted as a *provability* predicate. In the next step, he formalized the expected properties of this provability predicate by some axioms which have made the well-known modal system **S4**. Notice that in contrast with using a concrete interpretation of provability, he used a theory for formalizing this concept (**S4**). In fact, his system just characterizes the properties of our intuitive provability predicate by some formal system, and is totally silent about its real nature.

By this introduction, we are ready to give the definition of his interpretation. Consider the translation function $b : \mathcal{L} \to \mathcal{L}_\square$ as follows:
$\mathcal{L}$ and $\mathcal{L}_\square$ are the languages of **IPC** and **S4** respectively. [1]

(i) $p^b = \square p$ and $\perp^b = \square \perp$

(ii) $(A \wedge B)^b = A^b \wedge B^b$

(iii) $(A \vee B)^b = A^b \vee B^b$

(iv) $(A \to B)^b = \square(A^b \to B^b)$

(v) $(\neg A)^b = \square(A^b \to \square \perp)$

Now it is clear that A^b is the implicit BHK interpretation of A. In fact, the definition of b is the natural paraphrase of the original BHK interpretation in terms of the provability instead of the proofs.

Now, it is time to investigate the soundness-completeness property of the interpretation.

---

[1]In fact, our translation is different from the translation of the paper [6]. The differences are the following: $p^b = p$, $\perp^b = \perp$, $(A \to B)^b = \square A^b \to \square B^b$, and $(\neg A)^b = \neg \square A^b$. While, both of these two translations basically do the same task, we use the first one, because it is more compatible with our intuition of intuitionistic semantics and it is adequate for the systems weaker than **S4**.

2. Soundness and Completeness: Consider the following theorem:

Theorem 1.1. *For any proposition $A \in \mathcal{L}$, **IPC** $\vdash A$ iff **S4** $\vdash A^b$.*

Proof. For the complete investigation of this theorem and some related results[2] see [6]. $\qquad\square$

Now we have the system **S4** which formalizes what we expect from a provability predicate and based on the mentioned soundness-completeness result we can reduce the problem of finding a formalization of the implicit BHK interpretation to the problem of finding a provability interpretation for **S4**. Therefore, our task will be to find a concrete interpretation of this provability predicate (the connective box) in terms of the classical provability in classical theories. But, consider the fact that the problem of finding a provability interpretation for **S4** has its own importance itself, independent on its relation to the BHK interpretation.

The first attempt to find a concrete provability interpretation for **S4** was made by Gödel himself. In a very negative way, he showed that the natural expected interpretation of the provability predicate is not sound for **S4**. Let us explain his result in a more detailed way:
The most natural choice to interpret the box operator is the provability predicate of a formal theory [3]. Let T be a formal system; therefore, the meaning of $\Box A$ would be $\mathrm{Pr}_T(A)$ such that $\mathrm{Pr}_T(\cdot)$ is a provability predicate for T. (Notice that in this case we suppose our formal system, T, to be sufficiently strong to be able to formalize some parts of the meta-mathematics.) Now, consider the theorem $\Box\neg\Box \perp$ of **S4**. Its interpretation is $\mathrm{Pr}_T(\neg\mathrm{Pr}_T(\perp))$ and if it would be true we have $T \vdash \neg\mathrm{Pr}_T(\perp)$ which contradicts Gödel's second incompleteness theorem.
Therefore, we know that on the one hand, the natural way to formalize the concept of proof and provability in the BHK interpretation is to fix a formal system and interpret all the proofs as the proofs in that theory. And on the other hand, the logic **S4** is not sound with this natural interpretation. This is for the case of **S4**. However, we could claim that the natural formalization for the BHK interpretation is not sound as well. For instance, if you try to interpret the sentence $A \wedge (A \to B) \to B$ in intuitionistic logic, you find out that it is more or less the same as the modal formula $\Box(\Box p \to p)$ and you

---

[2]While this theorem is the heart of Gödel's work, he only stated it and left it without any proof. The soundness part is an easy consequence of induction on the length of the proof, but the completeness part was finally proved in 1947 by Tarski and McKinsey by the algebraic semantics for **S4**.

[3]The system T is formal iff the set of its consequences is recursively enumerable.

will encounter with the same problem in intuitionistic logic. In sum, we can say that the natural formalization of the BHK interpretation and also the natural interpretation of **S4** do not work. Based on these observations, we have an intuition why finding a formalization for the BHK interpretation is complicated and hard to grasp.

Now there is a natural question to ask. If the theory **S4** is intuitively valid and we know that we can not interpret the box as a provability predicate in some formal system, then, what could be a natural provability interpretation of **S4**? Unfortunately, despite a lot of attempts which have been made so far, this question remains open. For instance, Kripke [7] introduced a provability interpretation which is based on his Kripke models and just captures our provability intuition for formulas without nested modalities. Or in [3], Buss introduced the "pure provability" which has the same problem with the nested modalities. Actually, the only successful attempt to find a provability interpretation, is Artemov's "logic of proofs" which is based on the idea of introducing all explicit proofs, investigating the intended behavior of proofs in a theory (logic of proofs) and then interpreting the box as the existence of the proof. These explicitly mentioned proofs could empower us to avoid non-standard proofs which has the main role in Gödel's second incompleteness theorem and all counter-intuitive theorems in meta-mathematics which make some problems to interpret **S4**. In the section eight we will come back to Artemov's logic of proofs and we will investigate its advantages and disadvantages.

Now, as this long introduction shows, the main problem is to find a provability interpretation for the modal logic **S4** to formalize the BHK interpretation. In this paper, we will try to solve this problem and in the forthcoming part of the introduction we will sketch the idea of our semantics and our key results.

The Main Idea and the Main results

Why doesn't the mentioned natural proof interpretation work? The answer is the fact that this interpretation does not distinguish between languages and metalanguages. Let us illuminate this fact by an example. Suppose p is an atom. What should be an intended interpretation of p? p is an atomic sentence about the real world, it is just a description of the world and this description is in the first level. But how about $\Box p$? The intended interpretation of this formula is the provability of p in some theory. But, what is important here, is the level of the theory and the level of this sentence. Since

p is a fact about the real world, the theory in which p is proved, should be a first level theory, i.e. a theory about the world. However, the sentence $(\Box p)$ is not about the real world; it is about the provability and hence it should be characterized as a sentence in the second level. Therefore, the intended meaning of this second level sentence is $\Pr_{T_0}(p)$. Now let us ask about the interpretation of $\Box\Box p$. This is about the provability of the provability of p. The first box refers to a first level theory T_0. But the second box is about the provability of the provability, which has higher order, and it means the provability should be investigated in a second level theory like T_1. Now, the important thing is the fact that there is no reason to assume that $T_1 = T_0$. Actually, our experience in mathematical logic shows that it is genuinely important to distinguish the meta-theory and the object theory, and in most cases the power of the meta-theory should be more than the theory itself. For instance, Gödel's incompleteness theorems show that to answer a very basic meta-mathematical question about the system, i.e. its consistency, we need a more powerful meta-theory. Based on these investigations, the natural way to interpret boxes in a modal sentence is interpreting them in different theories with respect to the complexity of the occurrence of a box. To formalize this idea, we need two different objects. First, a model for the real world to interpret atoms as the facts about the world and second a hierarchy of theories which plays the role of the hierarchy of the meta-theories. Hence, the intended model would be $(M, \{T_n\}_{n=0}^{\infty})$ in which M is a classical model and T_n is a theory in the n-th level of the hierarchy. (We call these models, the provability models.) Moreover, we need a way of witnessing all boxes as the provability predicates of these theories in an appropriate way. This is the complex part of the formalization and we will talk about it in the next section. But for now, just think about the interpretation intuitively in the sense that any outer box should be interpreted as the provability predicate of a bigger theory. Therefore, our main result for modal logics is the following:

Theorem 1.2.   *(i) The logic **K4** is sound and complete with respect to the provability interpretation in all provability models.*

(ii) *The logic **KD4** is sound and complete with respect to the provability interpretation in consistent provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that T_n is consistent and $T_{n+1} \vdash \mathrm{Cons}(T_n)$.*

(iii) *The logic **S4** is sound and complete with respect to the provability interpretation in all reflexive provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that T_n is sound and $T_{n+1} \vdash \mathrm{Rfn}(T_n)$.*

(iv) *The logic **GL** is sound and complete with respect to the provability interpretation in all constant provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where*

for any n, M thinks that $T_n = T_0$.

(v) *The logic **GLS** is sound and complete with respect to the provability interpretation in all sound constant provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that T_n is sound and $T_n = T_0$.*

(vi) *The extensions of the logic **KD45** are not sound in any provability model.*

There are some remarks about this main theorem. First of all, it shows that the use of a hierarchy of meta-theories instead of just one theory to witness the box operators, could define a brand new framework to capture different modal logics in terms of the provability interpretations. In fact, it shows that modal logics could be seen as the formal theories to describe the relation between the real world and the theories in the hierarchy of meta-theories which we use; in other words, they are theories for the whole discourse of the provability. Moreover, in the case of the logics **K4**, **KD4** and **S4** it shows that they describe the relation of the model and meta-theories in a natural and expected way. For instance, in an informal reading of the axiom $\Box A \to A$ in **S4**, we mean that our proofs are sound. And this is exactly one of the conditions we put on the models to capture the logic **S4**. It is similar for all other axioms, logics and conditions in the aforementioned result.

Secondly, the result shows that if we restrict the whole hierarchy of meta-theories to just one theory, we could reconstruct Solovay's results for **GL** and **GLS**. Therefore, it shows that our provability interpretation is a generalization of Solovay's interpretation and our main result is a generalization of Solovay's results.

Now, if we combine this provability interpretations with Gödel translation, we will have different BHK interpretations with respect to different powers of meta-theories. Consider that by weak BHK interpretation, we informally mean the usual BHK interpretation without the consistency condition. This is the last condition in BHK interpretation which assumes that there is not any proof for $\bot$. We have:

Theorem 1.3. (i) *The logic **BPC** is sound and complete with respect to the BHK interpretation in all provability models.*

(ii) *The logic **BPC** $+ (\top \to \bot) \Rightarrow \bot$ is sound and complete with respect to the BHK interpretation in all consistent provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that T_n is consistent and $T_{n+1} \vdash \mathrm{Cons}(T_n)$.*

(iii) *The logic **MPC** is sound and complete with respect to the weak BHK interpretation in all reflexive provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that T_n is sound and $T_{n+1} \vdash \mathrm{Rfn}(T_n)$.*

(iv) *The logic **IPC** is sound and complete with respect to the BHK interpretation in all reflexive provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that T_n is sound and $T_{n+1} \vdash \mathrm{Rfn}(T_n)$.*

(v) *The logic **FPL** is sound and complete with respect to the BHK interpretation in all constant provability models, i.e. $(M, \{T_n\}_{n=0}^{\infty})$ where for any n, M thinks that $T_n = T_m$.*

(vi) *The logic **CPC** does not admit any BHK interpretations.*

There are some remarks about this result. First of all, it shows that there are different BHK interpretations instead of just one. This observation, somehow contradicts the folklore belief and it is surprising. The reason is that the BHK interpretation just defines the meaning of a connective in terms of the provability in different levels of meta-languages. But, it is silent about what kinds of commitments we impose on our meta-theories.

Therefore, we can impose different philosophical conditions on the behavior of meta-theories to capture different propositional logics, all of them valid under the BHK interpretation. For instance, we can choose the minimal possible commitment which means that there is not any non-trivial condition on the hierarchy of meta-theories. Then the BHK interpretation leads to the logic **BPC**. On the other hand, if we suppose that our meta-theories are strong enough to prove the reflection principle for lower theories and all the theories are sound, then the BHK interpretation leads to the logic **IPC**. This observation shows a key fact: There is a web of different intuitionistic logics according to the BHK interpretation; the logics **IPC** and **BPC** are just two examples of these intuitionistic logics and both of them are philosophically valid. In sum, we have to talk about *intuitionistic logics* instead of *the* intuitionistic logic.

Secondly, the result shows that this framework of the provability interpretation can capture different propositional logics and just like the case of modal logics, we are able to say that propositional logics are logics to describe the behavior of the real world and the hierarchy of meta-theories. This formalizes the intuitionist claim that the intuitionistic mathematics is a way to talk and only talk about proofs.

Thirdly, it is possible to define different kinds of Gödel's translation. Hence, it is possible to capture different propositional logics via these different translations. But it is important to consider that the translation we used in the

above result is the valid translation to formalize the BHK interpretation and those different kinds of translations may not be rooted in the usual BHK interpretation. However, they are still provability interpretations and could be useful.

2   Provability models

In this section we will introduce a provability model as a formalization of the intuitive combination of a model and a hierarchy of theories. Then, we will define the satisfaction relation between modal formulas and provability models. And as a conclusion, we will justify our notion of provability interpretation.

First of all we have a modal formula A, and we want to interpret any box in the formula as a provability predicate. This is the core idea of a provability interpretation for A. The second idea is what we talked about in the introduction. Intuitively, when you have two boxes in A such that one box is in the scope of the other box, our intuition forces us to accept that the outer box talks about the provability in the meta-theory while the inner box is just capturing the provability in the lower theories. Based on these two ideas, we can claim that the natural model for the provability interpretation of modal logics is a pair of one first order structure to interpret the atoms of the language, and a hierarchy of theories to play the role of a hierarchy of meta-theories. Moreover, we choose our structure and our theories as a model and theories for arithmetic, respectively, because in these theories we have a natural way of coding the language, the meta-language, the meta meta language and so on. Furthermore, we suppose that all of our theories include $I\Sigma_1$ to have enough power to formalize the basic meta-mathematics of the theories. And, for the same reason we assume $M \vDash I\Sigma_1$, because we want to have the true meta-mathematical properties obviously.

Definition 2.1. A provability model is a pair $(M, \{T_n\}_{n=0}^{\infty})$ where M is a model of $I\Sigma_1$ and $\{T_n\}_{n=0}^{\infty}$ is a hierarchy of arithmetical r.e. theories such that for any n, $I\Sigma_1 \subseteq T_n \subseteq T_{n+1}$ provably in $I\Sigma_1$.

Now we define an extension of a modal formula.

Definition 2.2. $E(A)$, the set of all extensions of A, is inductively defined as follows:

- If A is an atom, $E(A) = \{A\}$.

- If $A = B \circ C$, then $E(A) = \{D \circ E \mid D \in E(B) \text{ and } E \in E(C)\}$ for $\circ \in \{\wedge, \vee, \rightarrow\}$.

10

- If $A = \neg B$, then $E(A) = \{\neg D \mid D \in E(B)\}$.

- If $A = \Box B$, then $E(A) = \{\Box \bigvee_{i=1}^{k} D_i \mid \forall 1 \leq i \leq k, \ D_i \in E(B)\}$.

Moreover, if Γ is a sequence of modal formulas, by a sequence of extensions of Γ, we mean a sequence such that for any formula in Γ, it has at least one of its extensions and at most finitely many of them. We will denote these sets by $\bar{\Gamma}$.

Informally speaking, an extension of a formula A is a formula resulted by changing any formula after a box with a bunch of disjunctives of the extensions of the formula.

Example 2.3. For instance, the formula $\Box(\neg\Box(\Box p \vee \Box p) \vee \neg\Box\Box(p \vee p))$ is an extension of the formula $\Box\neg\Box\Box p$

So far, we have justified the definition 2.1. Now let us to investigate the intuitive meaning of the witnesses. We claim that a natural interpretation is based on the interpretion of the outer boxes as meta-theories of the inner boxes. For simplicity, we call this kind of interpretation as the ordered interpretation. Therefore, to have an ordered interpretation we need to interpret all of the boxes in A as the provability predicates of the theories in an ordered way. And, since for any theory we have a number which shows its layer in the hierarchy, it is enough to assign a natural number to a box. Consider that if we assign n to a box, the intended meaning is that the interpretation of that box is the provability predicate for the theory T_n. This role is played by the concept of witness. In fact, a witness is just an assignment for the boxes in an ordered way.

Notation 2.4. If w_is are sequences of the natural numbers, by $(w_1, w_2, \ldots, w_n)$ we mean the concatenation of w_is.

Definition 2.5. Let w be a sequence of natural numbers and A be a modal formula. Then the relation $w \Vdash A$, which means w is a witness for A, is inductively defined as follows:

- If A is an atom, $() \Vdash A$.

- If $A = B \circ C$, then $(w_1, w_2) \Vdash A$ if $w_1 \Vdash B$ and $w_2 \Vdash C$ for $\circ \in \{\wedge, \vee, \rightarrow\}$

- If $A = \neg B$, then $w \Vdash A$ if $w \Vdash B$.

- If $A = \Box B$, then $(n, w) \Vdash A$ if $w \Vdash B$ and $n > m$ for all m which are appeared in w.

11

Moreover, if Γ is a sequence of modal formulas, by a witness for Γ we mean, a sequence of witnesses, such that any witness w_i in the sequence, is a witness for A_i in Γ.

Informally, a witness for a formula A means a sequence of numbers which we assign to occurrences of the boxes in A such that the number for outer box is greater than all numbers of inner boxes. This condition, formalizes the idea that any outer box refers to the meta-theories in the hierarchy.

Example 2.6. For instance, $w = (n, m, k, r)$ is a witness for $\Box(p \to q) \vee \Box(\neg\Box p \to \Box q)$ if $m > k, r$.

The next definition is about evaluating a modal formula under an arithmetical substitution for atoms and a witness for the boxes in the formula.

Definition 2.7. Let w be a witness for A and σ be an arithmetical substitution which assigns an arithmetical sentence to a propositional variable. And also let $(M, \{T_n\}_{n=0}^\infty)$ be a provability model. By $A^\sigma(w)$ we mean an arithmetical sentence which is resulted by substituting the variables by σ and interpreting any box as the provability predicate of T_n if the correspondent number in the witness for this box was n. The interpretation of boolean connectives are themselves. Moreover, if Γ is a sequence of modal formulas A_i, and $w = (w_i)_i$ is its witness, by $\Gamma^\sigma(w)$ we mean the sequence of $A_i^\sigma(w_i)$.

Example 2.8. For the witness and the formula A of the last example, $A^\sigma(w)$ would be $\mathrm{Pr}_{T_n}(p^\sigma \to q^\sigma) \vee \mathrm{Pr}_{T_m}(\neg\mathrm{Pr}_{T_k}(p^\sigma) \to \mathrm{Pr}_{T_r}(q^\sigma))$.

Now, we are ready to introduce the concept of the satisfiability of a formula in a provability model.

Definition 2.9. A sequent $\Gamma \Rightarrow \Delta$ is true in $(M, \{T_n\}_{n=0}^\infty)$ when there are sequences of extensions of Γ and Δ, like $\bar{\Gamma}$ and $\bar{\Delta}$, and also witnesses u and v for $\bar{\Gamma}$ and $\bar{\Delta}$ respectively, such that for any arithmetical substitution σ, $M \models \bar{\Gamma}^\sigma(u) \Rightarrow \bar{\Delta}^\sigma(v)$. Moreover, we say a sequent $\Gamma \Rightarrow \Delta$ is true in a class of models C, when there are uniform sequences of extensions and witnesses for all models. In a more precise way, we write $C \vDash \Gamma \Rightarrow \Delta$, if there are sequences of extension $\bar{\Gamma}$ and $\bar{\Delta}$ and witnesses u and v such that for all arithmetical substitutions σ and all provability models $(M, \{T_n\}_{n=0}^\infty)$ in C, $M \models \bar{\Gamma}^\sigma(u) \Rightarrow \bar{\Delta}^\sigma(v)$.

Informally speaking, truth means the existence of extensions and witnesses such that the interpretation of a formula (or sequent) becomes true, independent of the use of the arithmetical substitutions.

Remark 2.10. Note that our definition of satisfiability allows us to use a disjunction of finitely many extensions of the formula instead of the original formula itself. In other words, if we want to show that $(M, \{T_n\}_{n=0}^\infty) \vDash A$, we could use finitely many extensions of A like $B_1, B_2, \ldots, B_k$ and find a witness for $\bigvee_{i=1}^k B_i$. The same is true for the sequents.

Let us illuminate the definition 2.9 with some examples.

Example 2.11. Let $(\mathbb{N}, \{T_n\}_{n=0}^\infty)$ be a provability model where $T_0 = \text{PA}$ and for any n, $T_{n+1} = T_n + \text{Rfn}(T_n)$. Based on the definition, this pair is obviously a provability model. Now, we want to show that the sentence $\Box(\Box A \rightarrow A)$ is true in the model. To do this, we need some extensions of the formula and a witness for them. For the extensions, just use the formula itself, and for a witness, first find a witness for A and call it w; if n is a number greater than all the numbers in w, then the sequence $(n+1, n, w, w)$ is a witness for $\Box(\Box A \rightarrow A)$. Now, for any arithmetical substitution σ, we have $\mathbb{N} \vDash \text{Pr}_{T_{n+1}}(\text{Pr}_{T_n}(A^\sigma(w)) \rightarrow A^\sigma(w))$ since the theory T_{n+1} can prove the reflection for T_n. As you can see, the idea of introducing a hierarchy to witness the boxes in modal sentences could kill the effect of Gödel's second incompleteness theorem.

Now, let us illuminate the importance of the extensions with an example. Consider the sentence $\neg\Box(\neg\Box A \wedge A)$. We want to show that this sentence is true in the above mentioned provability model. (Note that this formula is provable in **S4**.) Pick a witness w for the sentence A, a number n greater than all numbers in w and the formula itself as its extension. In this case we need two copies of the sentence, therefore we have to find a witness for $B = \neg\Box(\neg\Box A \wedge A) \vee \neg\Box(\neg\Box A \wedge A)$. It is easy to verify that the sequence $(n+2, n+1, w, w, n+1, n, w, w)$ is a witness for B. Now, for any arithmetical interpretation like σ, we have

$$\mathbb{N} \vDash \neg\text{Pr}_{n+2}(\neg\text{Pr}_{n+1}(A^\sigma(w)) \wedge A^\sigma(w)) \vee \neg\text{Pr}_{n+1}(\neg\text{Pr}_n(A^\sigma(w)) \wedge A^\sigma(w))$$

Because if we have both

$$\text{Pr}_{n+2}(\neg\text{Pr}_{n+1}(A^\sigma(w)) \wedge A^\sigma(w))$$

and

$$\text{Pr}_{n+1}(\neg\text{Pr}_n(A^\sigma(w)) \wedge A^\sigma(w))$$

then from the first part and the soundness of T_{n+2} we have $\neg\text{Pr}_{n+1}(A^\sigma(w))$ and from the second part and the fact that the provability predicate commutes with $\wedge$, we have $\text{Pr}_{n+1}(A^\sigma(w))$, which is a contradiction. Therefore, the sentence is true in $\mathbb{N}$. It is easy to see that if we want to show the truth

of the sentence $\Box(\neg\Box(\neg\Box A \wedge A))$, we should use $\Box B$ as an extension of the formula. This observation shows the importance of the extensions, but is it possible to avoid them?

Example 2.12. In this example we want to argue that some sentences do not have a witness in some provability models. Finding these kinds of examples is not a hard task to do. It is enough to think about formulas like p or $\Box p$. However, what we want to show here is finding an example to show the importance of the extensions in the definition. Think about the provability model of the last example and consider the formula $\neg\Box(\neg\Box p \wedge p)$. We showed that if we use two different copies of the formula, then the disjunction of those different copies have a witness in the provability model. Now, we want to show that if we just use one copy, it is impossible to witness the formula. Assume that $w = (n, m)$ is a witness for $\neg\Box(\neg\Box p \wedge p)$ in the above mentioned provability model. Then since w is a witness, we have $n > m$. On the other hand, we know that for any arithmetical substitution, we have $\mathbb{N} \vDash \neg\mathrm{Pr}_n(\neg\mathrm{Pr}_m(p^\sigma) \wedge p^\sigma)$. Use the arithmetical substitution which sends p to $\mathrm{Cons}(T_m)$. Therefore, we have

$$\mathbb{N} \vDash \neg\mathrm{Pr}_n(\neg\mathrm{Pr}_m(\mathrm{Cons}(T_m)) \wedge \mathrm{Cons}(T_m))$$

Based on the formalized Gödel's second incompleteness theorem

$$I\Sigma_1 \vdash \mathrm{Cons}(T_m) \rightarrow \neg\mathrm{Pr}_m(\mathrm{Cons}(T_m))$$

since $I\Sigma_1 \subseteq T_{m+1}$ and $T_{m+1} \vdash \mathrm{Cons}(T_m)$ we have

$$T_{m+1} \vdash \neg\mathrm{Pr}_m(\mathrm{Cons}(T_m))$$

hence $\mathbb{N} \vDash \mathrm{Pr}_{m+1}(\neg\mathrm{Pr}_m(\mathrm{Cons}(T_m)))$ and since T_{m+1} has the reflection principle for T_m, $\mathbb{N} \vDash \mathrm{Pr}_{m+1}(\mathrm{Cons}(T_m))$. Now since $n > m$ we have

$$\mathbb{N} \vDash \mathrm{Pr}_n(\neg\mathrm{Pr}_m(\mathrm{Cons}(T_m))) \wedge \mathrm{Pr}_n(\mathrm{Cons}(T_m))$$

which contradicts our assumption. As you can see, our provability interpretation is sensitive to the use of extensions and also to the numbers of copies of extensions. In the following discussion, we will show that this property is an inherent property of the informal intuition behind the modal formulas.

Discussion

One of the complexities of our provability interpretation is the use of extensions and in this discussion, we want to justify its role. But before that, we

need some observations. First of all, it seems that if we use the intuitive interpretation of the boxes as the provability predicates of different theories in the hierarchy of theories, meta-theories, meta-meta theories and so on, the natural provability interpretation will be the following:

A sentence A is true in a provability model $(M, \{T_n\}_{n=0}^{\infty})$, if there is a witness w for A such that for all σ, $M \models A^{\sigma}(w)$.

Which informally says that if you could witness the boxes in the formula A in the provability model, then it is true. Note that this definition is more simple than ours and does not use any kinds of extensions. Now, let us concentrate on **S4** as the theory for our intuitive provability, and temporarily use the above definition as the definition of the truth. To interpret all axioms of the system **S4**, it is easy to see that we need two natural conditions on our model. First of all T_{n+1} should be powerful enough to prove the reflection of the theory T_n and secondly, all T_ns should be sound with respect to our model M (This is what the nature of the provability in **S4** assumes; think about $\Box(\Box A \to A)$ and $\Box A \to A$ respectively.) Now, the sentence $\neg\Box(\neg\Box A \wedge A)$ is a theorem of **S4** and we expect that it should be true in any model with those two conditions. But in example 2.12 we showed that there is not any witness for the sentence and hence, with the definition above, the sentence is not true. Why does it happen? The answer is that it is because of the different roles of one occurrence of box in modal formulas. To illuminate this fact, let us investigate the intuitive proof of the sentence $\neg\Box(\neg\Box A \wedge A)$ in **S4**. The proof is a proof by contradiction. Assume $\Box(\neg\Box A \wedge A)$, then because all theorems are true (axiom **T**), we have $\neg\Box A \wedge A$ and hence $\neg\Box A$. On the other hand, since the provability commutes with the conjunction (a consequence of the axiom **K**), we have $\Box A$, which is a contradiction. Now consider the fact that the box in $\neg\Box A$ is inherited from the inner box in $\neg\Box A \wedge A$ and the box in $\Box A$ is inherited from the outer box in $\neg\Box A \wedge A$. Therefore, to reach the contradiction, we need these two boxes refer to one layer in the hierarchy of theories which is impossible because the inner one is the theory and the other is the meta-theory and it is impossible to have $T_{n+1} = T_n$, because T_{n+1} should prove the reflection for T_n.

What these investigations show, is actually the fact that one box in **S4** could have different roles. (In the above sentence, the outer box has two different roles, one as the meta-theory of the inner box and the other, as the theory itself.) Therefore, the natural way to interpret these boxes, is an approach which captures the different roles of a box at the same time, and this is not possible with the above simplified semantics; because it is obviously based on the assumption that any box has just one role which needs just one witness. Now, here is where we need extensions. In fact, the intended meaning of the

extensions, is using different copies of the formula in a disjunction and if you witness this disjunction, you have the power to witness one box in finitely different ways; this technique empowers us to capture different roles of one box. (See example 2.11 to find out how this technique works.)

Now there is another question to ask. Why do we need this kind of iterative extension method and why is just the simple disjunction of the formula not enough? The answer is that for any fixed role available for one box, it is also possible to have different roles for inner boxes. Therefore, after any box you need a new disjunction. (Think about the sentence $\Box(\neg\Box(\neg\Box A \wedge A))$.) This is just what we call extensions.

As a conclusion for this discussion, let us compare our situation here in modal logic with the first order logic. In the first order logic, if we have a theorem of the form $\forall x \exists y A(x, y)$ where $A(x, y)$ is quantifier-free and if we want to witness y, the Herbrand's theorem gives the answer; we can witness y by terms in our language. However, we know that one term is not enough. The reason is simple. The existentially quantified y could have different values (roles) and these different values (roles) can be captured by a disjunction of sentences $A(x, t(x))$ for some finite possible set of terms $t(x)$. Now, the situation in modal logic is the same. We read boxes as the existence of theories and we want to witness them. Since there are different roles for any box, we need a disjunction to capture these different roles. In other words, we could interpret the extensions as some kind of Herbrandization of the modal formulas.

3   The Logic K4

Intuitively, the logic **K4** is sound with respect to all kinds of provability interpretations. The reason is very simple. **K4** has two important modal axioms; the axiom **K** which means that the provability predicate is closed under modus ponens, and the axioms **4** which means that the provability of a sentence is also provable. The first axiom is a very easy fact and all strong enough meta-theories can prove it. On the other hand, if we have the minimum power in our meta-theory (Σ_1-completeness), the axiom **4** would be also easily proved. Now, consider the fact that these axioms are not only true but also provable and it justifies the use of the necessitation rule. Hence, **K4** is valid in all provability interpretations.

Now, in this section we want to formalize this intuitive argument and show that the logic **K4** is sound and also strongly complete with respect to the class of all provability models. First, the soundness theorem. If we show the class of all provability models by **PrM**, then we have:

Theorem 3.1. *(Soundness) If* $\Gamma \vdash_{\mathbf{K4}} A$ *then* $\mathbf{PrM} \vDash \Gamma \Rightarrow A$.

Proof. To prove the soundness theorem for $\mathbf{K4}$, we will use the cut-free sequent calculus for $\mathbf{K4}$ i.e. $\mathbf{GK4}$ (see [8]). To simplify the proof, we use the following conventions: Firstly, if Φ and Ψ are sequences of arithmetical sentences and T is an arithmetical theory, by $T \vdash \Phi \Rightarrow \Psi$, we mean $T \vdash \bigwedge \Phi \rightarrow \bigvee \Psi$. Secondly, without loss of generality, we assume that the main formulas in all of the rules, except the exchange rule, are just the rightmost formulas in the sequent. We just use this assumption for the sake of brevity and clearance of the proof.

Now, we want to prove the following claim by induction on the length of the proof in $\mathbf{GK4}$.

Claim. If $\Gamma \Rightarrow \Delta$ is provable in $\mathbf{GK4}$, then there are sequences of extensions like $\bar{\Gamma}$ and $\bar{\Delta}$ and also witnesses w_1 and w_2 for $\bar{\Gamma}$ and $\bar{\Delta}$ respectively, such that for any provability model $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ, $I\Sigma_1 \vdash \bar{\Gamma}^{\sigma}(w_1) \Rightarrow \bar{\Delta}^{\sigma}(w_2)$.

1. The case of axioms and structural rules. For the axiom $A \Rightarrow A$, it is enough to use A as its extension in both sides and just an arbitrary witness for A in both sides, again.

For the exchange rule, just use the same extensions and witnesses after the application of the correspondent exchange.

For the weakening rule, if we prove $\Gamma, A \Rightarrow \Delta$ from $\Gamma \Rightarrow \Delta$, by IH, we could find extensions $\bar{\Gamma}$, $\bar{\Delta}$ and witnesses w_1 and w_2. Pick an arbitrary witness for A like w. Now for $\Gamma, A \Rightarrow \Delta$, use the sequences $\bar{\Gamma}, A$ and $\bar{\Delta}$, and for the witnesses use (w_1, w) and w_2. It is easy to show that $I\Sigma_1 \vdash \bar{\Gamma}^{\sigma}(w_1), A^{\sigma}(w) \Rightarrow \bar{\Delta}^{\sigma}(w_2)$. The case for the right weakening is the same.

For the contraction rule, if we prove $\Gamma, A \Rightarrow \Delta$ from $\Gamma, A, A \Rightarrow \Delta$, then by IH, there are sequences of extensions $\{\bar{\Gamma}, \{\bar{A}_{i1}\}_{i=0}^{r}, \{\bar{A}_{j2}\}_{j=0}^{s}\}$ and Δ and also witnesses $w_1 = (u, (v_{i1})_{i=0}^{r}, (v_{j2})_{j=0}^{s})$ and w_2. Now for the sequent $\Gamma, A \Rightarrow \Delta$, use the sequences of extensions $\{\bar{\Gamma}, \{\bar{A}_{i1}\}_{i=0}^{r}, \{\bar{A}_{j2}\}_{j=0}^{s}\}$ and $\bar{\Delta}$ and for the witnesses just use the same witnesses. In this case, because of the use of a finite set of different extensions instead of just one extension, we can say that the semantics absorbs the contraction rule. The case for the right contraction is the same.

17

2. The case of propositional rules. In this case we just prove the case that the last rule is $R\wedge$; the other rules are similar and the argument is the same. If $\Gamma_1, \Gamma_2 \Rightarrow \Delta_1, \Delta_2, A\wedge B$, is proved from $\Gamma_1 \Rightarrow \Delta_1, A$ and $\Gamma_2 \Rightarrow \Delta_2, B$ then by IH we have the sequences of extensions $\bar{\Gamma}_1, \{\bar{\Delta}_1, \{\bar{A}_i\}_{i=0}^r\}, \bar{\Gamma}_2, \{\bar{\Delta}_2, \{\bar{B}_j\}_{j=0}^s\}$ and witnesses w_1 and $w_2 = (u, (x_i)_{i=0}^r)$ and $w_1', w_2' = (u', (y_j)_{j=0}^s)$. Now, for the sequent $\Gamma_1, \Gamma_2 \Rightarrow \Delta_1, \Delta_2, A\wedge B$ use the sequences of extensions $\{\bar{\Gamma}_1, \bar{\Gamma}_2\}$, $\{\bar{\Delta}_1, \bar{\Delta}_2, \{\bar{A}_i \wedge \bar{B}_j\}_{i=0,j=0}^{i=r,j=s}\}$ and witnesses (w_1, w_1'), $(u, u', ((x_i, y_j))_{i=0,j=0}^{i=r,j=s})$.

3. The case of modal rules. If $\Box\Gamma \Rightarrow \Box A$ is proved from $\Gamma, \Box\Gamma \Rightarrow A$, then by IH, we have the sequences of extensions $\{\bar{\Gamma}_1, \overline{\Box\Gamma}_2\}$ and $\{A_i\}_{i=0}^r$ and witnesses $w_1 = ((u_j)_{j=0}^s, (v_k)_{k=0}^t)$ and $w_2 = (x_i)_{i=0}^r$ where u_j is a witness for the jth formula in $\bar{\Gamma}_1$ and v_k is a witness for the kth formula in $\overline{\Box\Gamma}_2$. Pick number n greater than all the numbers in w_1 and w_2. Now, for the sequent $\Box\Gamma \Rightarrow \Box A$ use the sequences of extensions $\{\overline{\Box\Gamma}_1, \overline{\Box\Gamma}_2\}$ and $\Box\bigvee_{i=0}^r A_i$ and for the witnesses use $((n, u_j)_{j=0}^s, (v_k)_{k=0}^t)$ and $(n, (x_i)_{i=0}^r)$. By IH, we know that for any arithmetical substitution σ,

$$I\Sigma_1 \vdash \bigwedge_{j=0}^s \bar{\Gamma}_1^\sigma(u_j) \wedge \bigwedge_{k=0}^t \overline{\Box\Gamma}_2^\sigma(v_k) \to \bigvee_{i=0}^r A_i^\sigma(x_i).$$

Since $I\Sigma_1 \subseteq T_n$, we have

$$T_n \vdash \bigwedge_{j=0}^s \bar{\Gamma}_1^\sigma(u_j) \wedge \bigwedge_{k=0}^t \overline{\Box\Gamma}_2^\sigma(v_k) \to \bigvee_{i=0}^r A_i^\sigma(x_i).$$

Therefore, by Σ_1-completeness in $I\Sigma_1$ we have

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge_{j=0}^s (\bar{\Gamma}_1^\sigma(u_j) \wedge \bigwedge_{k=0}^t (\overline{\Box\Gamma}_2^\sigma(v_k))) \to \bigvee_{i=0}^r A_i^\sigma(x_i)),$$

hence

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge_{j=0}^s \bar{\Gamma}_1^\sigma(u_j)) \wedge \mathrm{Pr}_n(\bigwedge_{k=0}^t \overline{\Box\Gamma}_2^\sigma(v_k)) \to \mathrm{Pr}_n(\bigvee_{i=0}^r A_i^\sigma(x_i)).$$

Now, by formalized Σ_1-completeness of T_n in $I\Sigma_1$ we have

$$I\Sigma_1 \vdash \bigwedge_{k=0}^t \overline{\Box\Gamma}_2^\sigma(v_k) \to \mathrm{Pr}_n(\bigwedge_{k=0}^t \overline{\Box\Gamma}_2^\sigma(v_k))$$

and hence

$$I\Sigma_1 \vdash \bigwedge_{j=0}^s \mathrm{Pr}_n(\bar{\Gamma}_1^\sigma(u_j)) \wedge \bigwedge_{k=0}^t \overline{\Box\Gamma}_2^\sigma(v_k) \to \mathrm{Pr}_n(\bigvee_{i=0}^r A_i^\sigma(x_i)),$$

which is what we wanted to prove and it completes the proof of the claim. For the proof of the soundness theorem, if $\Gamma \vdash_{\mathbf{K4}} A$ then there exist a finite set $\Delta \subseteq \Gamma$ such that $\Delta \vdash_{\mathbf{K4}} A$. Therefore, $\mathbf{GK4} \vdash \Delta \Rightarrow A$. By claim there are some extensions of Δ and A like $\bar{\Delta}$ and $\{A_i\}_{i=0}^r$ and witnesses u and $\{w_i\}_{i=0}^r$ such that for any arithmetical substitution σ, we have $I\Sigma_1 \vdash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(w_i)$. Since $M \vDash I\Sigma_1$, we have $M \vDash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(w_i)$. Now, pick $\bar{\Gamma}$ the same as Γ after changing the part of Δ to $\bar{\Delta}$. Moreover, choose v as a witness for $\bar{\Gamma}$ as an arbitrary extension of u to $\bar{\Gamma}$. Hence, $M \vDash \bar{\Gamma}^\sigma(v) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(w_i)$ which completes the proof of the soundness. $\square$

For the completeness theorem, the idea is reducing the completeness of $\mathbf{K4}$ to the completeness of $\mathbf{GL}$ which is the well-known Solovay's theorem. (See [10].) To do that, we need some kind of translation from $\mathbf{K4}$ to $\mathbf{GL}$ which could transfer the provability behavior of $\mathbf{K4}$ to the provability behavior of $\mathbf{GL}$.

Definition 3.2. Let A be a modal formula with k boxes. And $Q = \{q_i\}_{i=0}^\infty$ is a sequence of atoms which are not used in A. Then, a translation t based on Q for the modal sentence A, is a sequence of k numbers which assigns natural numbers to boxes in A such that the number assigned to the outer box is greater than all the numbers for the inner boxes. And A^t is defined as follows:

(i) If A is an atom, $A^t = A$.

(ii) $(B \circ C)^t = B^t \circ C^t$ for all $\circ \in \{\wedge, \vee, \rightarrow\}$

(iii) $(\neg B)^t = \neg B^t$.

(iv) $(\Box B)^t = \Box(\bigwedge_{i=0}^n q_i \rightarrow B^t)$ where n is the assigned number to the box in t.

Informally, if we interpret a box as the provability predicate for the theory S, then the translation t is just changing the provability predicate of the theory S to the provability predicate of the theory $S + \{q_0, \ldots, q_n\}$ where n is the assignment of the box under t. For instance, if $t = (1, 2, 1)$ and $A = \Box p \rightarrow \Box\Box p$, then A^t will be the following modal formula:

$$\Box(q_0 \wedge q_1 \rightarrow p) \rightarrow \Box(q_0 \wedge q_1 \wedge q_2 \rightarrow \Box(q_0 \wedge q_1 \rightarrow p))$$

Now, we want to show that this translation is complete, i.e.

Theorem 3.3. *If* $\mathbf{GL} \vdash A^t$ *for some translation* t, *then* $\mathbf{K4} \vdash A$.

The natural proof should be based on a technique of the transformation of transitive Kripke models to conversely well-founded transitive Kripke models, which is implemented by the following lemma.

Lemma 3.4. *Let (K, R, V) be a finite transitive Kripke tree with clusters, A be a modal formula and t a translation. Then there is a finite transitive irreflexive Kripke model (K', R', V') such that for any node $k \in K$, there is a node $k' \in K'$ such that if $k \vDash A$ then $k' \vDash A^t$.*

Proof. First of all, for all subformulas B of A, define the complexity of B, $C(B)$, as follows: If B is box-free, define $C(B) = -1$. Otherwise, define $C(B)$ as the maximum assigned number in B by t. Moreover, suppose that $C(A) = n$. Now, to simplify the proof, let us make some conventions. We will use I for clusters and for any $k \in K$, by $I(k)$ we mean the cluster of k. By a path, we mean a sequence of nodes in K, and if all the nodes of the path p belong to the cluster I, we write $p \subset I$. Moreover, we write $p \prec p'$, when p is a proper first segment of p'. Finally, by $e(p)$ we mean the rightmost element of p, or in other words, the end of p.

Now, for any cluster I define $X(I)$ as follows: If I consists of one irreflexive node k, $X(I) = \{k\}$ and if I consists of some finite reflexive nodes, define $X(I)$ as the subset of all paths $p \subset I$ with length less than or equal to $n + 2$. The idea is simple. We want to transform a transitive model to a nonreflexive transitive model. To accomplish this, we will unwind the reflexive clusters by some paths of nodes in that cluster and we will use q's to have a way to refer to a copy of the node instead of itself, when we check the truth of the modal formulas.

Now, define $K' = \bigcup_I X(I)$ and $R' = R_1 \cup R_2$ where

$$R_1 = \bigcup_{(k,l) \in R, I(k) \neq I(l)} \{(a, b) \mid a \in X(I(k)) \text{ and } b \in X(I(l))\}$$

and

$$R_2 = \bigcup_I \{((p, p') \mid p \prec p'; p, p' \subset I\}.$$

And finally, define

$$V'(r) = \{p \in K' \mid e(p) \in V(r)\} \cup \{k \mid k \in V(r) \text{ and } k \text{ is irreflexive}\}$$

for all atoms r in A, and

$$V(q_i) = \{k \mid k \text{ is irreflexive}\} \cup \{p \mid |p| \leq n + 2 - i\}.$$

Informally speaking, K' is just the set K when you change the reflexive cluster I with all paths of length less that or equal to $n+2$ of nodes in I; R' and V' are the natural relation and valuation induced by R and V, respectively and q_i is true in all irreflexive nodes and also in all paths of nodes in reflexive clusters with length bounded by $n+2-i$. Now, we want to prove the following two claims.

Claim.1. The model (K', R', V') is a finite transitive irreflexive Kripke model.

The finiteness follows from the definition. For the transitivity, suppose that $a, b, c \in K'$ and $(a, b) \in R'$ and $(b, c) \in R'$. Then, there are two cases. The first case is when a and b come from the same cluster. Hence, by definition, this cluster should be a reflexive cluster. Therefore, a and b are paths in this cluster and $a \prec b$. Now, if c comes also from this cluster, we will have $b \prec c$ and since $\prec$ is transitive, we have $a \prec c$ and hence $(a, c) \in R'$. But, if c comes from another cluster, then the cluster of c should be above the cluster of b and hence it is also above the cluster of a which is the same as b's and then by definition we have $(a, c) \in R'$.
The proof of the second case, which is when a and b come from different clusters, is similar to the proof of the first case.

For the irreflexivity, suppose $(a, a) \in R'$. If a is an irreflexive node in K, then it is impossible, by the definition of R', to have $(a, a) \in R'$. If a comes from a reflexive cluster, then again by the definition of R', the path a should be a proper segment of itself which is impossible.

Now, we express the second claim.

Claim.2. For all subformulas of A like B, if $k \vDash B$, then

$$\begin{cases} \forall p, |p| \le n+1 - C(B) \wedge e(p) = k, \ p \vDash B^t & \text{if } k \text{ is reflexive.} \\ k \vDash B^t & \text{if } k \text{ is irreflexive.} \end{cases}$$

and if $k \nvDash B$ then

$$\begin{cases} \forall p, |p| \le n+1 - C(B) \wedge e(p) = k, \ p \nvDash B^t & \text{if } k \text{ is reflexive.} \\ k \nvDash B^t & \text{if } k \text{ is irreflexive.} \end{cases}$$

To prove the claim, we use induction on B.

1. Atomic case. If B is an atom, the claim easily follows from the definition of V'.

2. If $B = C \wedge D$ and $k \vDash C \wedge D$ then $k \vDash C$ and $k \vDash D$. If k is irreflexive, then by IH, the claim holds. If k is reflexive, then by IH, for all p such that $|p| \leq n + 1 - C(C)$ and $e(p) = k$, we have $p \vDash C^t$. And also for all p such that $|p| \leq n + 1 - C(D)$ and $e(p) = k$, we have $p \vDash D^t$, and since $C(C \wedge D) = max\{C(C), C(D)\}$, then for all p such that $|p| \leq n + 1 - C(C \wedge D)$ and $e(p) = k$, we have $p \vDash C^t \wedge D^t$.
If $k \nvDash C \wedge D$, then $k \nvDash C$ or $k \nvDash D$. W.l.o.g. assume $k \nvDash C$. If k is irreflexive, the claim is obvious. If k is reflexive, then by IH, for all p such that $|p| \leq n + 1 - C(B)$ and $e(p) = k$ we have $p \nvDash C^t$, and again since $C(C \wedge D) = max\{C(C), C(D)\}$ we have $\forall p, |p| \leq n + 1 - C(B \wedge D) \wedge e(p) = k,\ p \nvDash (C \wedge D)^t$.

3. If $B = \neg C$, then for irreflexive k, the claim is obvious from IH. Now, if k is reflexive and $k \vDash \neg C$, then $k \nvDash C$, and by IH, $\forall p, |p| \leq n + 1 - C(C)\ p \nvDash C^t$. Therefore, $\forall p, |p| \leq n + 1 - C(C)\ p \vDash \neg C^t$ and since $C(C) = C(\neg C)$ we have what we wanted. The other case is the dual of the first case.

4. The case for disjunction and implication is the same as the cases for conjunction and negation and we omit them here.

5. The modal case. This is the most important and the most complex part of the proof.

5.1. If $B = \Box C$ and $k \vDash \Box C$ then for all l which $(k, l) \in R$, $l \vDash C$. Now, define $C(B) = m$.
5.1.1. If k is irreflexive, we know that the nodes above k in K' are of two forms. The l's which are irreflexive and $(k, l) \in R$ or the p's where p comes from a cluster I above k and $e(p) = l$. For the first kind of nodes, by IH we know that $l \vDash C^t$, therefore $l \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. If we were in the second case, we know that $l \vDash C$ and again by IH, for all p such that $|p| \leq n + 1 - C(C)$ and $e(p) = l$, we have $p \vDash C^t$. Therefore, for all $p, |p| \leq n + 1 - C(C)$ we have $p \vDash C^t$ and hence $p \vDash \bigwedge_{i=0}^{k} q_i \to C^t$. If $|p| > n + 1 - C(C)$, since $C(C) < C(B) = m$, we have $|p| > n + 2 - m$, and then by the definition of the valuation we know that $p \nvDash q_m$ and hence $p \nvDash \bigwedge_{i=0}^{m} q_i$ and thus $p \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. Therefore, for all p above k, we have $p \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. Now since for all nodes above k, $\bigwedge_{i=0}^{m} q_i \to C^t$ is true, we have $k \vDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$ which means $k \vDash (\Box C)^t$.

5.1.2. If k is reflexive from the cluster I, pick p such that $|p| \leq n + 1 - m$.

We want to show that $p \vDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. We know that all nodes above p are of the form irreflexive l's or $p' \subset J$ where J is a cluster above I or $p' \subset I$ where $p \prec p'$. For the first and second kinds, by a similar proof of 5.1.1, we can show that $l \vDash \bigwedge_{i=0}^{m} q_i \to C^t$ and $p' \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. For the third case, if $|p'| > n + 2 - m$, then $p' \nvDash q_m$ and hence $p' \nvDash \bigwedge_{i=0}^{m} q_i$ and thus $p' \vDash \bigwedge_{i=0}^{m} q_i \to C^t$. If $|p'| \le n + 2 - m$ then since $C(C) \le m - 1$ we have $|p'| \le n + 1 - C(C)$. On the other hand, $k \vDash \Box C$, hence all nodes in I satisfies C, and specially we have $e(p') \vDash C$, by IH, and by the fact that $|p'| \le n + 1 - C(C)$, we have $p' \vDash C^t$ and therefore $\bigwedge_{i=0}^{m} q_i \to C^t$. Now we proved that for all nodes above p, we have $\bigwedge_{i=0}^{m} q_i \to C^t$ hence $p \vDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$ which is what we wanted.

5.2. If $B = \Box C$ and $k \nvDash \Box C$, then there is a node like l such that $l \nvDash C$. Define $C(B) = m$.

5.2.1. If k is irreflexive, we want to show that $k \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. Consider that since $(k, l) \in R$, and k is irreflexive, then $l \ne k$ and it belongs to a cluster above k. Now, if l is irrefelexive then by IH, $l \nvDash C^t$ and also since it is irreflexive, for all i, $l \vDash q_i$; hence $l \nvDash \bigwedge_{i=0}^{m} q_i \to C^t$ since $l \ne k$ and $(k, l) \in R$, $(k, l) \in R'$. Therefore, $k \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. If l is a reflexive node of the cluster I, then define $p \subset I$ as a path such that $|p| = n + 2 - m$ and $e(p) = l$. Since $C(C) \le m - 1$ then $|p| \le n + 1 - C(C)$. By IH, $p \nvDash C^t$. (Consider that m is the complexity of a boxed formula and therefore $m \ge 0$, hence $n + 2 - m \le n + 2$ and it means such a p exists.). Moreover, we know that $p \vDash \bigwedge_{i=0}^{m} q_i$ since $|p| \le n + 2 - i$ for all $i \le m$, therefore, $p \nvDash \bigwedge_{i=0}^{m} q_i \to C^t$. Now since the cluster of k and the cluster of l are different and $(k, l) \in R$, then $(k, p) \in R'$ and it means that $k \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$.

5.2.2. Consider the case that k is reflexive. In this case, if l belongs to a cluster above k, then the proof is the same as 5.2.1. If the cluster of l and k are the same (say I), we have the following construction: Pick p such that $e(p) = k$ and $|p| \le n + 1 - m$. We want to show that $p \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$. Pick $p' \subset I$ such that $e(p') = l$, $p \prec p'$ and $|p'| = n + 2 - m$. (It is enough to extend p to a path with the ending l and the length $n + 2 - m$. Note that $n + 2 - m > n + 1 - m$, which guarantee the existence of the extension with the ending l possibly different from k. Moreover, this length is less that $n + 2$ and therefore p' exists in our model as a path). We know that $C(C) \le m - 1$, hence $|p'| \le n + 1 - C(C)$. By IH, $p' \nvDash C^t$. On the other hand, $p \vDash \bigwedge_{i=0}^{m} q_i$ since $|p| \le n + 2 - i$ for all $i \le m$, therefore, $p \nvDash \bigwedge_{i=0}^{m} q_i \to C^t$. Now, since $p \prec p'$, we can conclude that $p \nvDash \Box(\bigwedge_{i=0}^{m} q_i \to C^t)$.

Now, the lemmas are obvious by the claim 2. For B in the claim, choose

23

A itself, then if $k \vDash A$ and k is irreflexive , then $k \vDash A^t$. But if k is reflexive, pick $p = k$ as a path with length one. Hence $|p| = 1 \leq n + 1 - C(A)$, since $C(A) = n$ and therefore, $p \vDash A^t$. Therefore, for any $k \vDash A$ there is a node in K' like k' such that $k' \vDash A^t$. $\qquad\square$

Now for the proof of theorem 3.3 we have:

Proof. If $\mathbf{K4} \nvdash A$, then there is a finite transitive Kripke tree with clusters like (K, R, V) and a node k such that $k \vDash \neg A$. If we apply lemma 3.4 for $\neg A$, we can construct a finite transitive irreflexive Kripke model (K', R', V') and a node k' such that $k' \nvDash \neg A^t$. But (K', R', V') is a model of $\mathbf{GL}$ and $\mathbf{GL} \vdash A^t$. A contradiction. Hence $\mathbf{K4} \vdash A$. $\qquad\square$

Now, based on the completeness of the translations, which we have introduced, we are able to prove the completeness theorem. But, since we want to establish a more powerful completeness result, i.e. the strong completeness, we need one more lemma.

Lemma 3.5. *There is a hierarchy of theories like $\{T_n\}_{n=0}^{\infty}$ such that for any n, $I\Sigma_1 \subseteq T_n$ and $T_n \subseteq T_{n+1}$ provably in $I\Sigma_1$ and also an arithmetical substitution $*$ such that for any modal formula A, if there exists a witness w for A such that $(M, \{T_n\}_{n=0}^{\infty}) \vDash A^*(w)$ for all $M \vDash I\Sigma_1$, then $\mathbf{K4} \vdash A$.*

Proof. Add infinitely many new atoms like $Q = \{q_n\}_{n=0}^{\infty}$ to the language of modal logics. And apply all axioms and rules of the logic $\mathbf{K4}$ to the new language to construct a new system $\mathbf{K4(Q)}$ and also do the same thing for the logic $\mathbf{GL}$ to construct $\mathbf{GL(Q)}$. Now, pick the substitution $*$ as the uniform substitution of Solovay's theorem (see [2]). It simply says that for any A, $I\Sigma_1 \vdash A^*$ iff $\mathbf{GL(Q)} \vdash A$, where A^* means the combination of substituting any atom p with p^* and interpreting all boxes as the provability predicate of $I\Sigma_1$. Now, for any n, define $T_n = I\Sigma_1 + \{q_i^*\}_{i=0}^{n}$. We claim that this $*$ and this hierarchy $\{T_n\}_{n=0}^{\infty}$ works for the claim of the lemma. First of all, it is easy to show that the hierarchy has the claimed conditions. Secondly, we have $M \vDash A^*(w)$ for all $M \vDash I\Sigma_1$. Therefore, $I\Sigma_1 \vdash A^*(w)$. Use q_i's in the translations from $\mathbf{K4}$ to $\mathbf{GL}$. Since the interpretation of a box in any formula like $\Box D$ with witness m is $\mathrm{Pr}_{T_m}(D)$, and it is provably equivalent to $\mathrm{Pr}_{I\Sigma_1}(\bigwedge_{i=0}^{m} q_i \to D)$, it is easy to see that there is a translation t, such that $I\Sigma_1 \vdash A^*(w) \leftrightarrow (A^t)^*$. (In fact t equals to the witness w.) Therefore, $I\Sigma_1 \vdash (A^t)^*$, by the uniform version of Solovay's theorem, $\mathbf{GL(Q)} \vdash A^t$, and by theorem 3.3, $\mathbf{K4(Q)} \vdash A$. It means that there exists a proof for A in $\mathbf{K4(Q)}$. Now, since A does not have any $q_i \in Q$, it is enough to put $q_i = \top$ everywhere in the proof to find a proof for A in $\mathbf{K4}$. $\qquad\square$

Now we want to prove the strong completeness theorem.

Theorem 3.6. *(Strong Completeness) If* $\mathbf{PrM} \vDash \Gamma \Rightarrow A$*, then* $\Gamma \vdash_{\mathbf{K4}} A$*.*

Proof. We know that there are the sequence of extensions $\bar{\Gamma}$, and extensions $B_1, \ldots, B_k$ of A and witnesses u for $\bar{\Gamma}$, and $w_1, \ldots, w_k$ for $B_1, \ldots, B_k$ such that for all provability models and all arithmetical substitution σ,

$$M \vDash \bar{\Gamma}^\sigma(u) \Rightarrow \{B_i^\sigma(w_i)\}_{i=0}^k.$$

Now, pick the hierarchy and $*$ from lemma 3.5. Then for all $M \vDash I\Sigma_1$,

$$M \vDash \bar{\Gamma}^*(u) \Rightarrow \{B_i^*(w_i)\}_{i=0}^k.$$

Hence

$$I\Sigma_1 + \bar{\Gamma}^*(u) \vdash \bigvee_{i=0}^k B_i^*(w_i).$$

Therefore there is a finite subset of $\bar{\Gamma}$ like Δ and a subset of witnesses v from u, such that

$$I\Sigma_1 + \Delta^*(v) \vdash \bigvee_{i=0}^k B_i^*(w_i).$$

Hence, for all $M \vDash I\Sigma_1$, we have

$$M \vDash \bigwedge \Delta^*(v) \to \bigvee_{i=0}^k B_i^*(w_i).$$

By lemma 3.5, $\mathbf{K4} \vdash \bigwedge \Delta \to \bigvee_{i=0}^k B_i$, which means $\bar{\Gamma} \vdash_{\mathbf{K4}} \bigvee_{i=0}^k B_i$. Finally, since in the presence of the axiom $\mathbf{K}$, all extensions of a formula are equivalent to itself, $\Gamma \vdash_{\mathbf{K4}} A$. $\qquad \square$

4 The Logic KD4

The logic $\mathbf{KD4}$ is a modal logic resulted by adding the axiom $\mathbf{D} : \Box A \to \neg\Box\neg A$ to $\mathbf{K4}$. It is easy to see that we can change this axiomatization with adding the axiom $\neg\Box \perp$ instead of $\mathbf{D}$. Now, intuitively, if we have the consistency of theories and also they are provable in their meta-theories, then the axioms of $\mathbf{KD4}$ are valid. (Since we have the neccesitation rule, the sentence $\Box\neg\Box \perp$ is also provable and this is why we need the consistency statements to be provable, as well.) The formalization of these models is exactly what we will call consistent provability models and we will show that the logic $\mathbf{KD4}$ is sound and strongly complete with respect to these models.

Definition 4.1. A provability model $(M, \{T_n\}_{n=0}^{\infty})$ is called consistent if for any n, M thinks that T_n is consistent and $T_{n+1} \vdash \mathrm{Cons}(T_n)$, i.e. $M \vDash \mathrm{Cons}(T_n)$ and $M \vDash \mathrm{Pr}_{T_{n+1}}(\mathrm{Cons}(T_n))$. Moreover, the class of all consistent provability models will be shown by **Cons**.

Now, let us prove the soundness theorem.

Theorem 4.2. *(Soundness)* If $\Gamma \vdash_{\mathbf{KD4}} A$, then $\mathbf{Cons} \vDash \Gamma \Rightarrow A$.

Proof. We use the soundness theorem for **K4**. If $\Gamma \vdash_{\mathbf{KD4}} A$, then

$$\Gamma + \Box\neg\Box \bot \wedge \neg\Box \bot \vdash_{\mathbf{K4}} A.$$

Now, based on the soundness of **K4**, there are sequences of extensions of $\Gamma + \Box\neg\Box \bot \wedge \neg\Box \bot$ and A like $\bar{\Gamma} + \{\Box(\bigvee_{j=0}^{s_i} \neg\Box \bot) \wedge \neg\Box \bot\}_{i \in I}$ and $\{A_k\}_{k=0}^{t}$ and witnesses u, $(n_i, (m_{ij})_{j=0}^{s_i}, k_i)$ and w_k such that for any provability model like $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ,

$$M \vDash \bar{\Gamma}^{\sigma}(u) + \{\mathrm{Pr}_{n_i}(\bigvee_{j=1}^{s_i} \neg\mathrm{Pr}_{m_{ij}}(\bot)) \wedge \neg\mathrm{Pr}_{k_i}(\bot)\}_{i \in I} \Rightarrow \bigvee_{k=0}^{t} A_k^{\sigma}(w_k))$$

Now, if we apply this fact on the consistent provability models, since $n_i > m_{ij}$ and for any n, $M \vDash \mathrm{Pr}_{n+1}(\neg\mathrm{Pr}_n(\bot))$, we have $M \vDash \mathrm{Pr}_{n_i}(\neg\mathrm{Pr}_{m_{ij}}(\bot))$ for all $i \leq r$ and $j \leq s_i$. Moreover, since for any n, $M \vDash \neg\mathrm{Pr}_n(\bot)$, we have $M \vDash \neg\mathrm{Pr}_{k_i}(\bot)$. Therefore, for any consistent provability model like $(M, \{T_n\}_{n=0}^{\infty})$ we have

$$M \vDash \bar{\Gamma}^{\sigma}(u) \Rightarrow \bigvee_{k=0}^{t} A_k^{\sigma}(w_k)$$

which completes the proof of the soundness for **KD4**. $\qquad\square$

For the completeness theorem, the idea is reducing the completeness of **KD4** to the completeness of **K4** which was proved in the previous section.

Theorem 4.3. *(Strong Completeness)* If $\mathbf{Cons} \vDash \Gamma \Rightarrow A$, then $\Gamma \vdash_{\mathbf{KD4}} A$.

Proof. We know that there are the multiset $\bar{\Gamma}$, and extensions $B_1, \ldots, B_k$ of A and witnesses u for $\bar{\Gamma}$, and $w_1, \ldots, w_k$ for $B_1, \ldots, B_k$ such that for any consistent provability model and any arithmetical substitution σ,

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \bar{\Gamma}^{\sigma}(u) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Now, define Δ as a sequence which consists of an infinite number of the formula $\Box\neg\Box \bot$ and also an infinite number of the formula $\neg\Box \bot$. We claim

that $\Gamma, \Delta \Rightarrow A$ is true in the class **PrM**. For the extensions, use the same extensions for Γ and A, and also use Δ itself, as its sequence of extensions. For witnesses, use u, w_i's and for Δ, for any number n, use $(n+1, n)$ for one of the formulas $\Box \neg \Box \perp$ and n for one of the formulas $\neg \Box \perp$. Call this witness v. Now, let $(M, \{T_n\}_{n=0}^{\infty})$ be an arbitrary provability model. We claim that

$$M \vDash \bar{\Gamma}^{\sigma}(u), \Delta^{\sigma}(v) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Because when $M \vDash \bar{\Gamma}^{\sigma}(u), \Delta^{\sigma}(v)$ then $M \vDash \Delta^{\sigma}(v)$ which means for any n,

$$M \vDash \mathrm{Pr}_{n+1}(\neg \mathrm{Pr}_n(\perp)),$$

and

$$M \vDash \neg \mathrm{Pr}_n(\perp).$$

Therefore, $(M, \{T_n\}_{n=0}^{\infty})$ is a consistent provability model and since $M \vDash \bar{\Gamma}^{\sigma}(u)$ we have,

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \bigvee_{i=0}^{k} B_i^{\sigma}(w_i).$$

Therefore, for all provability models and all σ, we have

$$M \vDash \bar{\Gamma}^{\sigma}(u), \Delta^{\sigma}(v) \Rightarrow \{B_i^{\sigma}(w_i)\}_{i=0}^{k}.$$

Hence, by the strong completeness of **K4**, we have $\Gamma, \Delta \vdash_{\mathbf{K4}} A$ and since all formulas in Δ are provable in **KD4**, we have $\Gamma \vdash_{\mathbf{KD4}} A$. □

Remark 4.4. Note that the truth of a formula in a class of provability models means the existence of a *uniform* sequence of extensions and also a *uniform* witness for it. In other words, we have a fix sequence of natural numbers which works for all provability models in the class. Therefore, we could claim that sentences just describe the behavior of the natural numbers instead of some actual theories. What does it mean? It means that sentences do not describe the behavior of a concrete specific provability model, but instead, they talk about the roles of these ingredients in the structure (provability model) which are encoded by the natural numbers. Informally speaking, sentences just transcend the actual theories to the their abstract roles in the structure of a provability model. (As an example, think about how the cardinal numbers transcend the concept of cardinality from the actual sets.) For instance, in the case of the logic **KD4**, it describes the relation between a meta-theory T_{n+1} and its theory T_n which is the condition that the meta-theory is powerful enough to show the consistency of the theory. This is not about some actual theories which we use; it is about the power of the meta-theory in comparison to its theory. In other words, **KD4** describes the abstract condition of *consistency* and *provability of consistency*. This fact is true in all soundness-completeness results we propose in this paper.

5 The Logic S4

Intuitively, if we have the property that all theories are sound and the soundness of theories are also provable in their meta-theories, all axioms of **S4**, would be valid. The formalization of these models is exactly what we will call the reflexive provability models. In fact, we will show that the logic **S4** is sound and also strongly complete with respect to the class of all reflexive provability models.

Definition 5.1. A provability model $(M, \{T_n\}_{n=0}^{\infty})$ is reflexive if for any n, M thinks that T_n is sound and $T_{n+1} \vdash \mathrm{Rfn}(T_n)$, i.e. $M \vDash \mathrm{Pr}_{T_n}(A) \to A$ and $M \vDash \mathrm{Pr}_{T_{n+1}}(\mathrm{Pr}_{T_n}(A) \to A)$ for any sentence A. Moreover, the class of all reflexive provability models will be shown by **Ref**.

Now let us prove the soundness theorem.

Theorem 5.2. *(Soundness) If $\Gamma \vdash_{\mathbf{S4}} A$, then $\mathbf{Ref} \vDash \Gamma \Rightarrow A$.*

Proof. To prove the soundness theorem, we will use the cut-free sequent calculus for **S4**, i.e. **GS4** (see [8]). And, we will use the conventions of theorem 3.1. Now, we want to prove the following claim:

Claim. If $\Gamma \Rightarrow \Delta$ is provable in **GS4**, then there are sequences of extensions like $\bar{\Gamma}$ and $\bar{\Delta}$ and also witnesses w_1 and w_2 for $\bar{\Gamma}$ and $\bar{\Delta}$, respectively and a number n greater than all the numbers in w_1 and w_2, such that for any reflexive provability model $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ, $T_n \vdash \bar{\Gamma}^\sigma(w_1) \Rightarrow \bar{\Delta}^\sigma(w_2)$ is true in M. We will call the number n as the context number.

The proof of the claim is by induction on the length of the proof of $\Gamma \Rightarrow \Delta$ and the proof for the non-modal cases are exactly like the proof of theorem 3.1. But the difference is just the presence of the context number n here. To find this number in all non-modal cases, if the case is the axiom case, any number works; for contraction and exchange, just use the same number in the induction hypothesis. For weakening, use the successor of the maximum of the context number of the induction hypothesis and the arbitrary chosen witness for the weakening formula. For the other cases, it is enough to use the maximum numbers of the induction hypothesis. Now we want to prove the case of the modal rules.

1. If $\Gamma, \Box A \Rightarrow \Delta$ is proved from $\Gamma, A \Rightarrow \Delta$, then by IH, we can find sequences of extensions $\{\bar{\Gamma}, \{A_i\}_{i=0}^r\}$, $\bar{\Delta}$ and witnesses $w_1 = (u, (x_i)_{i=0}^r)$

and w_2 and the context number n. Now for the sequent $\Gamma, \Box A \Rightarrow \Delta$, use the sequences of extensions $\{\bar{\Gamma}, \{\Box A_i\}_{i=0}^r\}$, $\bar{\Delta}$ and for the witnesses use $(u, ((n, x_i))_{i=0}^r)$, w_2 and for the context number use $n + 1$. By IH, we know that for all reflexive provability models and all arithmetical substitution σ, M thinks

$$T_n \vdash \bar{\Gamma}^\sigma(w_1), \{A_i^\sigma(x_i)\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2).$$

Now we claim that there is a proof, formalizable in $I\Sigma_1$, for the following statement: If $T_n \subseteq T_{n+1}$, $T_{n+1} \vdash \mathrm{Pr}_n(A_i^\sigma(x_i)) \to A_i^\sigma(x_i)$ for all $q \le i \le r$ and

$$T_n \vdash \bar{\Gamma}^\sigma(w_1), \{A_i^\sigma(x_i)\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2)$$

then

$$T_{n+1} \vdash \bar{\Gamma}^\sigma(w_1), \{\mathrm{Pr}_n(A_i^\sigma(x_i))\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2).$$

The proof is simple. We have $T_n \subseteq T_{n+1}$ and $T_{n+1} \vdash \mathrm{Pr}_n(A_i^\sigma(x_i)) \to A_i^\sigma(x_i)$. Therefore,

$$T_{n+1} \vdash \bar{\Gamma}^\sigma(w_1), \{\mathrm{Pr}_n(A_i^\sigma(x_i))\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2).$$

The proof just uses the fact that all first order tautologies are provable and Pr is closed under modus ponens and all of these properties are provable in $I\Sigma_1$. Now since $M \vDash I\Sigma_1$, M thinks that this implication is true. On the other hand both of premises are true in M, because of IH and the condition of being a reflexive provability model. Therefore, M thinks

$$T_{n+1} \vdash \bar{\Gamma}^\sigma(w_1), \{\mathrm{Pr}_n(A_i^\sigma(x_i))\}_{i=0}^r \Rightarrow \bar{\Delta}^\sigma(w_2),$$

which completes the proof.

2. If $\Box\Gamma \Rightarrow \Box A$ is proved from $\Box\Gamma \Rightarrow A$, then by IH we have sequences of extensions like $\overline{\Box\Gamma}$ and some extensions $\{A_i\}_{i=0}^r$ and witnesses w_1 and $(x_i)_{i=0}^r$ and a context number n such that for all arithmetical substitutions like σ, M thinks

$$T_n \vdash \overline{\Box\Gamma^\sigma(w_1)} \Rightarrow \{A_i^\sigma(x_i)\}_{i=0}^r.$$

Now, for the sequent $\Box\Gamma \Rightarrow \Box A$, use the extension $\overline{\Box\Gamma}$ and $\Box(\bigvee_{i=0}^r A_i)$, and the witnesses w_1 and $(n, (x_i)_{i=0}^r)$ and the context number $n + 1$.
Based on the Σ_1-completeness available in M, M thinks

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge \overline{\Box\Gamma^\sigma(w_1)} \to \bigvee_{i=0}^r (A_i^\sigma(x_i))).$$

Because the provability predicate commutes with the implications provably in $I\Sigma_1$, we have this property in M, hence

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigwedge \overline{\Box\Gamma^\sigma(w_1)}) \to \mathrm{Pr}_n(\bigvee_{i=0}^r (A_i^\sigma(x_i)))$$

is true in M. Now again by Σ_1-completeness, we have

$$I\Sigma_1 \vdash \bigwedge(\overline{\Box\Gamma^\sigma(w_1)}) \to \mathrm{Pr}_n(\bigvee_{i=0}^r (A_i^\sigma(x_i)))$$

true in M. And finally since T_{n+1} is an extension of $I\Sigma_1$ provably in $I\Sigma_1$, we have the inclusion in M, hence

$$T_{n+1} \vdash \bigwedge(\overline{\Box\Gamma^\sigma(w_1)}) \to \mathrm{Pr}_n(\bigvee_{i=0}^r (A_i^\sigma(x_i)))$$

is true in M which completes the proof of the claim.

Now for the proof of the soundness theorem, if $\Gamma \vdash_{\mathbf{S4}} A$ then there exists a finite subset Δ of Γ such that $\Delta \vdash_{\mathbf{S4}} A$. Then $\mathbf{GS4} \vdash \Delta \Rightarrow A$, then by the claim, there are sequences of extensions like $\bar{\Delta}$ and $\{A_i\}_{i=0}^r$ and the witnesses u and $(x_i)_{i=0}^r$ and a context number n such that for all reflexive provability models like $(M, \{T_n\}_{n=0}^\infty)$ and all arithmetical substitution σ, we have $T_n \vdash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(x_i)$ in M. Therefore, by soundness of T_n in M, we have $M \vDash \bar{\Delta}^\sigma(u) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(x_i)$. Now, define $\bar{\Gamma}$ as the sequence of extensions of Γ by using Γ and changing the subset Δ to $\bar{\Delta}$ and also use any arbitrary witnesses to extend u to a witness for $\bar{\Gamma}$. Call this new witness v. We have

$$M \vDash \bar{\Gamma}^\sigma(v) \Rightarrow \bigvee_{i=0}^r A_i^\sigma(x_i)$$

which is what we wanted to prove. $\qquad\square$

For the completeness theorem, the idea is the same as the idea of the original proof of Solovay's theorem. We will modify the technique of encoding Kripke models in arithmetic. In this case, we need to encode transitive reflexive trees with clusters. Therefore we have two tasks. Firstly, finding a method to encode the clusters and secondly, modifying Solovay's construction to work with reflexive trees instead of irreflexive ones.

Lemma 5.3. *Let m be a natural number and $\{T_n\}_{n=0}^N$ be an increasing hierarchy of theories such that $I\Sigma_1 \subseteq T_0$, and for any n, $T_{n+1} \vdash \mathrm{Rfn}(T_n)$. Therefore, there are arithmetical sentences A_1, A_2, ..., A_m such that:*

(i) *For any i and j, if $i \neq j$ then $I\Sigma_1 \vdash A_i \wedge A_j \to \bot$*

(ii) *$I\Sigma_1 \vdash \bigvee_{i=1}^m A_i$*

(iii) *For any $n \leq N$, and any $i \leq m$, $T_{n+1} \vdash \neg\mathrm{Pr}_{T_n}(\neg A_i)$*

(*iv*) *If we also assume that all theories in the hierarchy are consistent, then for any $n \leq N$ and any $i \leq m$, $\mathbb{N} \vDash \neg \mathrm{Pr}_{T_n}(\neg A_i)$ and $\mathbb{N} \vDash A_m$.*

Proof. First of all, we want to prove the following claim:

Claim. For any increasing reflexive hierarchy like $\{T_n\}_{n=0}^N$ and any natural number $1 \leq p$, there is another increasing hierarchy like $\{T_n'\}_{n=0}^{Np}$ such that for any $n \leq N$, $T_{np}' = T_n$ and for any $i \leq Np - 1$, $T_{i+1}' \vdash \mathrm{Cons}(T_i')$. Moreover, if all of the theories in the T hierarchy are consistent, all of the theories in the T' hierarchy will be consistent, as well.

To prove the claim, define T_i' as follows: For $i = np$, define $T_i' = T_n$, then for the any $np \leq i < (n+1)p - 1$ define T_{i+1}' inductively as the theory $T_i' + \mathrm{Cons}(T_i')$. First of all, we want to show that for any $np \leq i < (n+1)p-1$, $T_{i+1}' \subseteq T_{(n+1)p}'$ and also $T_{(n+1)p}'$ proves the reflection principle for T_{i+1}'. The proof is based on the induction on i. If $i = np$, we know that $T_{(n+1)p}'$ proves the consistency for T_{np}', hence $T_{np+1}' \subseteq T_{(n+1)p}'$. Moreover, since $T_{(n+1)p}' \vdash \mathrm{Cons}(T_{np}')$, it is easy to check that $T_{(n+1)p}'$ can prove the reflection principle for $T_{np+1}' = T_{np}' + \mathrm{Cons}(T_{np}')$. Now, suppose that we have the claim for i, and we want to prove it for $i + 1$. By IH, $T_{(n+1)p}'$ proves the reflection principle for T_i', hence it proves the consistency of T_i' and hence $T_{i+1}' \subseteq T_{(n+1)p}'$. Now, again, it is easy to show that since $T_{(n+1)p}' \vdash \mathrm{Cons}(T_i')$, $T_{(n+1)p}'$ also proves the reflection principle for $T_{i+1}' = T_i' + \mathrm{Cons}(T_i')$.

Now, we claim that for any i, $T_i' \subseteq T_{i+1}'$ and T_{i+1}' proves the consistency of T_i'. The proof is based on two different cases of the definition of T_{i+1}'. If we are in the first case, then $i + 1 = (n + 1)p$ for some n. Then by what we proved so far, the claim is obvious. If we are in the second case, then $T_{i+1}' = T_i' + \mathrm{Cons}(T_i')$, and hence the claim is again obvious from the definition.

Moreover, if the first hierarchy is consistent, then since all T_i''s are subtheories of $T_{Np}' = T_N$, the second hierarchy is consistent, as well.

Now, we want to prove the lemma. If $m = 1$, pick $A_1 = (0 = 0)$; then it is easy to verify that this sentence satisfies the conditions of the lemma. The reason is that T_{n+1} proves the consistency of T_n and hence $T_{n+1} \vdash \neg \mathrm{Pr}_n(0 \neq 0)$. Moreover, if all theories are consistent, then $\neg A_1$ is not provable in T_n.

Now assume that $m > 1$ and use the hierarchy T from the assumption of the lemma, and also use the aforementioned construction to construct the hierarchy T', for $p = 2m$. We want to define the sentences A_i based on this

31

new hierarchy. Define

$$B_r = \bigvee_{k=1}^{N} (\text{Cons}(T'_{2km-2r}) \wedge \neg\text{Cons}(T'_{2km-2r+1}))$$

for $1 \leq r \leq m-1$. Now, define $A_1 = B_1$ and $A_r = \bigwedge_{i=1}^{r-1} \neg B_i \wedge B_r$ for $2 \leq r \leq m-1$ and $A_m = \bigwedge_{i=1}^{m-1} \neg B_i$. We claim that these A_i's have the property in the lemma. First of all, because of the form of A_i's, it is obvious that any two different A_i and A_j are contradictory and also $\bigvee_{r=1}^{m} A_r$. In fact, these claims are first order tautologies and hence they are provable in $I\Sigma_1$. Now, we want to show that

$$T'_{2(n+1)m} \vdash \neg\text{Pr}_{T'_{2nm}}(\neg A_r)$$

For now, assume $r \neq 1, m$. Let us argue in $I\Sigma_1$. If $\neg A_r$ is provable in T'_{2nm}, then by definition $\bigvee_{i=1}^{r-1} B_i \vee \neg B_r$ is provable in T'_{2nm}. Now, from B_t, $t \leq r-1$, we could conclude

$$\bigvee_{I} (\text{Cons}(T'_{2km-2t})) \vee \bigvee_{J} (\neg\text{Cons}(T'_{2km-2t+1}))$$

where $I = \{k \mid 2km - 2t + 1 \geq 2nm + 1\}$ and $J = \{k \mid 2km - 2t + 1 < 2nm\}$. First of all, we know that T'_{2nm} proves $\text{Cons}(T'_{2km-2t+1})$ if $k \in J$. The reason is that if $k \in J$, then $2km - 2t + 1 < 2nm$ and since the consistency of any theory is provable in the higher theory in T' hierarchy, we can prove the consistency of $T'_{2km-2t+1}$ in T'_{2nm}. Therefore, we can conclude that the following is provable in T'_{2nm}.

$$\bigvee_{I} (\text{Cons}(T'_{2km-2t})).$$

On the other hand, we know that if $k \in I$, then $k \geq n+1$ because $2km - 2t + 1 \geq 2nm + 1$ is impossible when $k \leq n$. Therefore, $2km - 2t \geq 2(n+1)m - 2t$. Moreover, $2(n+1)m - 2t \geq 2(n+1)m - 2(r-1)$ since $t \leq r-1$, and since the hierarchy is increasing, $\text{Cons}(T'_{2km-2t})$ implies $\text{Cons}(T'_{2(n+1)m-2(r-1)})$. Hence, B_t implies $\text{Cons}(T'_{2(n+1)m-2(r-1)})$. Furthermore, from

$$\neg B_r = \bigwedge_{k=1}^{N} (\text{Cons}(T'_{2km-2r}) \to \text{Cons}(T'_{2km-2r+1}))$$

we conclude

$$\text{Cons}(T'_{2(n+1)m-2r}) \to \text{Cons}(T'_{2(n+1)m-2r+1}).$$

Therefore, we have

$$T'_{2nm} \vdash (\mathrm{Cons}(T'_{2(n+1)m-2r}) \to \mathrm{Cons}(T'_{2(n+1)m-2r+1})) \vee \mathrm{Cons}(T'_{2(n+1)m-2(r-1)}).$$

Hence

$$T'_{2nm} + \mathrm{Cons}(T'_{2(n+1)m-2r}) \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1}) \vee \mathrm{Cons}(T'_{2(n+1)m-2(r-1)}).$$

But we have $2(n+1)m - 2r + 1 \le 2(n+1)m - 2(r-1)$; therefore

$$T'_{2nm} \vdash \mathrm{Cons}(T'_{2(n+1)m-2(r-1)}) \to \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

And hence

$$T'_{2nm} + \mathrm{Cons}(T'_{2(n+1)m-2r}) \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

Since $r \le m$, we have $2(n+1)m - 2r + 1 \ge 2nm$, therefore we have

$$T'_{2(n+1)m-2r+1} \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1}).$$

Note that all the parts of this argument is formalizable in $I\Sigma_1$. Now, for the first time we want to use $T'_{2(n+1)m}$ to reach the contradiction. Since $1 \le r$, then $2(n+1)m - 2r + 1 < 2(n+1)m$, hence the consistency of $T'_{2(n+1)m-2r+1}$ is provable in $T'_{2(n+1)m}$. Therefore, since we are arguing in $T'_{2(n+1)m}$, we have the consistency of $T'_{2(n+1)m-2r+1}$. On the other hand, we showed

$$\mathrm{Pr}_{T'_{2(n+1)m-2r+1}}(\mathrm{Cons}(T'_{2(n+1)m-2r+1})).$$

By the formalized version of the second incompleteness theorem in $I\Sigma_1$, we know that if a theory proves its own consistency it is inconsistent; hence $T'_{2(n+1)m-2r+1}$ is inconsistent. A contradiction. Therefore, $T'_{2(n+1)m}$ could show that $\neg A_r$ is not provable in T'_{2nm}.

Note that the proof uses the form of $\neg A_r$ which has some positive B_t's and one negative B_r. But Now if we are in the cases $r = 1$ or $r = m$, then $\neg A_r$ has just positive B_t's or just negative B_t's. In these cases it is enough to use the part of the proof which investigates the corresponding B_t's. Again argue in $I\Sigma_1$. For the case, $r = 1$, if T'_{2nm} proves $\neg A_1$, then T'_{2nm} proves $\neg B_1$. Therefore,

$$T'_{2nm} \vdash \bigwedge_{k=1}^{N} (\mathrm{Cons}(T'_{2km-2}) \to \mathrm{Cons}(T'_{2km-1})).$$

Hence

$$T'_{2nm} \vdash (\mathrm{Cons}(T'_{2(n+1)m-2}) \to \mathrm{Cons}(T'_{2(n+1)m-1})).$$

33

Since $m \geq 1$, we have $2(n+1)m - 1 \geq 2nm$ and hence

$$T'_{2(n+1)m-1} \vdash (\mathrm{Cons}(T'_{2(n+1)m-2}) \to \mathrm{Cons}(T'_{2(n+1)m-1}))$$

and then since $2(n+1)m - 1 > 2(n+1)m - 2$, we have

$$T'_{2(n+1)m-1} \vdash \mathrm{Cons}(T'_{2(n+1)m-1}).$$

Now argue in $T'_{2(n+1)m}$. We have the consistency of $T'_{2(n+1)m-1}$. On the other hand, $T'_{2(n+1)m-1}$ proves its own consistency, hence by the formalized second incompleteness theorem, it should be inconsistent. A Contradiction. Therefore, $T'_{2(n+1)m}$ proves that $\neg A_1$ is not provable in T'_{2nm}.

For the proof of the case $r = m$, use the idea of I and J for positive B_t's. It is enough to use I and J, to show that if $\neg A_m$ is provable in T'_{2nm}, then $\mathrm{Cons}(T'_{2(n+1)m-2(m-1)})$ will be provable in $T'_{2(n+1)m-2(m-1)}$. After that, reaching a contradiction is the same as for the other cases.

Now, since $T'_{2nm} = T_n$, we have a proof for the part (iii). For (iv), if the hierarchy T is consistent, then the hierarchy T' is also consistent and hence if $\neg A_r$ is provable in T'_{2nm} then we have

$$T'_{2(n+1)m-2r+1} \vdash \mathrm{Cons}(T'_{2(n+1)m-2r+1})$$

for cases $1 < r < m$, and

$$T'_{2(n+1)m-1} \vdash \mathrm{Cons}(T'_{2(n+1)m-1})$$

for $r = 1$, and

$$T'_{2(n+1)m-2(m-1)} \vdash \mathrm{Cons}(T'_{2(n+1)m-2(m-1)})$$

for $r = m$. consider that the arguments for these statements are formalizable in $I\Sigma_1$ and hence they are true. For $1 < r < m$, by the second incompleteness theorem, $T'_{2(n+1)m-2r+1}$ should be inconsistent. A contradiction. Therefore, T'_{2nm} can not prove $\neg A_r$ and hence $T_n \nvdash \neg A_r$. The cases $r = 1, m$ are similar. For the second part of (iv), note that we know $A_m = \bigwedge_{r=1}^{m} \neg B_r$. Now, we want to show that all B_r's are false. We have

$$B_r = \bigvee_{k=1}^{N} (\mathrm{Cons}(T'_{2km-2r}) \wedge \neg\mathrm{Cons}(T'_{2km-2r+1}))$$

and since the whole T' hierarchy is consistent, all statements $(\mathrm{Cons}(T'_{2km-2r}) \wedge \neg\mathrm{Cons}(T'_{2km-2r+1}))$ are false and hence B_r is false. Then $\neg B_r$ is true and hence A_m is true. $\qquad\square$

Lemma 5.4. *Let (K, R) be a finite reflexive transitive tree with clusters and let k be one of the nodes in the root cluster. Moreover, let $(\mathbb{N}, \{T_n\}_{n=0}^N)$ be a reflexive provability model. Then there exists a set of arithmetical sentences $\{S_i\}_{i \in K}$ such that*

(i) *If $i \neq j$, $T_0 \vdash S_i \rightarrow \neg S_j$.*

(ii) *$T_{n+1} \vdash S_i \rightarrow \mathrm{Pr}_n(\bigvee_{(i,j) \in R} S_j)$.*

(iii) *If $(i, j) \in R$ then $T_{n+1} \vdash S_i \rightarrow \neg \mathrm{Pr}_n(\neg S_j)$.*

(iv) *$\mathbb{N} \vDash S_k$.*

Proof. Like Solovay's proof of the completeness of **GL**, define a recursive function $h : \mathbb{N} \rightarrow K$ such that

$$h(0) = k \text{ and } h(x+1) = \begin{cases} j & \text{if } (i,j) \in R \quad \text{and} \quad \mathrm{Prf}_N(x, \neg S_j) \\ h(x) & \text{otherwise} \end{cases}$$

where $S_j = P_{I(j)} \wedge A_j \wedge j = j$ and $P_{I(j)} = \exists y \forall x \geq y \ h(x) \in I(j)$ in which $I(j)$ means the cluster of j. Moreover, A_j's are the sentences constructed in lemma 5.3 for $m = Card(I(j))$ and the hierarchy $\{T_n + P_{I(j)}\}_{n=0}^N$. In addition, we choose A_k as the sentence A_m from lemma 5.3. By these sentences, we mean the sentences from the proof of lemma 5.3, and not what the lemma claims. The reason is that we have to be sure that these sentences are definable from the code of the function h which has not been defined yet. The reason is the following:

The function h should be defined based on the classical circular argument based on the fixed point lemma in $I\Sigma_1$. The important part is that the A_j's constructed in lemma 5.3 are arithmetical formulas based on the code of $P_{I(j)}$, which makes the whole circular argument possible. It is provable in $I\Sigma_1$ that h is a function. (Note that we put $j = j$ in the definition of S_j to make sure that there is at most one j such that x would be a proof for $\neg S_j$ and it makes h a function.) It is also provable that h eventually stops in some cluster and since h is a function, this cluster is unique. The existence of such cluster is an obvious application of the fact that h is an increasing function and the tree is finite. Note that all of these facts are provable in $I\Sigma_1$. Now, to prove (i), consider two cases. If i and j belong to different clusters, then $P_{I(i)}$ and $P_{I(j)}$ are contradictory based on what we claimed about the uniqueness of the limit cluster. This contradiction is also provable in $I\Sigma_1$ and hence in T_0. Now if i and j belong to the same cluster, then by lemma 5.3, we know that A_i and A_j are contradictory, provable in $I\Sigma_1$, and hence we reach a contradiction

35

for $S_i \wedge S_j$ in T_0. For (ii), we argue in T_{n+1}. If we have S_i, then we have $P_{I(i)}$ and there exists x such that $h(x) \in I(i)$. Since this formula is Σ_1, by Σ_1-completeness we have $\mathrm{Pr}_n(h(x) \in I(i))$. Moreover, h is provably increasing in $I\Sigma_1$ and hence in T_n, and also provably in $I\Sigma_1$ we know that h eventually stops in some cluster, i.e. $\mathrm{Pr}_n(\bigvee_J P_J)$. But we have $\mathrm{Pr}_n(h(x) \in I(i))$. Therefore, the limit should be above i which means $\mathrm{Pr}_n(\bigvee_{(i,j)\in R} P_{I(j)})$. On the other hand, by lemma 5.3 we know that $I\Sigma_1 \vdash \bigvee_{i\in I} A_i$, and we can conclude that $\mathrm{Pr}_n(\bigvee_{(i,j)\in R} P_{I(j)} \wedge A_j)$, hence $\bigvee_{(i,j)\in R} S_j$.

For (iii), we will argue in T_{n+1} and the proof is by contradiction. If we have S_i and $\mathrm{Pr}_n(\neg S_j)$ for some j which $(i, j) \in R$, then there are two possibilities. First, when the clusters of i and j are different. We have $S_i = P_{I(i)} \wedge A_i$, hence we have $P_{I(i)}$ which means that there is some number z, such that for all $y \geq z$, $h(y) \in I(i)$. Moreover, we know that $\mathrm{Pr}_n(\neg S_j)$ and since $T_n \subseteq T_N$, we have $\mathrm{Pr}_N(\neg S_j)$. Therefore, there exists some x such that $\mathrm{Prf}_N(x, \neg S_j)$. It is easy to see that we can pick $x \geq z$. Hence, we can conclude that $h(x+1) \in I(i)$. Now, since $(i, j) \in R$, j is above all nodes in $I(i)$ and $\mathrm{Prf}_N(x, \neg S_j)$, hence $h(x+1) = j$. But $h(x+1)$ should belong to $I(i)$ and $j \notin I(i)$; a contradiction. Therefore, $\neg\mathrm{Pr}_n(\neg S_j)$.
Now assume that the cluster of i and j is I. Then the statement $S_i \to \mathrm{Pr}_n(\neg S_j)$ is equivalent to

$$P_I \wedge A_i \to \mathrm{Pr}_n(P_I \to \neg A_j).$$

Since $\{T_n\}_{n=0}^N$ is a reflexive hierarchy, the hierarchy $\{T_n + P_I\}_{n=0}^N$ is also reflexive. Moreover, A_t's are constructed for this hierarchy, hence by lemma 5.3, we know that

$$T_{n+1} + P_I \vdash \neg\mathrm{Pr}_{T_n+P_I}(\neg A_j)$$

which proves what we wanted.

For (iv), since h eventually stops in some cluster, there is a cluster I, such that $\mathbb{N} \vDash P_I$. If $I \neq I(k)$, since $h(0) = k$, there should be some first element like x, such that $h(x) \in I$. Assume $h(x) = i$. Since $x \neq 0$, and $h(x) \neq h(x-1)$, we have $\mathrm{Prf}_N(x-1, \neg S_i)$ and hence, $\mathrm{Pr}_N(P_I \to \neg A_i)$. By lemma 5.3, the theory $T_N + P_I$ should be inconsistent, and therefore we have $T_N \vdash \neg P_I$. On the other hand, the theory T_N is sound, hence $\mathbb{N} \vDash \neg P_I$ which contradicts to our assumption. Hence, $I = I(k)$ and therefore, $\mathbb{N} \vDash P_{I(k)}$. On the other hand, $T_N + P_{I(k)}$ is consistent because it is sound, and consequently by lemma 5.3, A_k which was chosen to be the A_m from the lemma, is true; hence $S_k = P_{I(k)} \wedge A_k$ is true. $\qquad \square$

The following lemma, uses the previous lemma to transfer the truth from a Kripke model to a reflexive provability model.

Lemma 5.5. *Assume the conditions of lemma 5.4 and let $\{S_i\}_{i \in K}$ be defined as in that lemma. Define σ as the arithmetical substitution which sends the atom p to $\bigvee_{i \Vdash p} S_i$. Now, for any $i \in K$, any modal formula A and any witness w for A with elements less than N, we have:*

$$\begin{cases} T_{max(w)+1} \vdash S_i \to A^\sigma(w) & \textit{if } i \vDash A \\ T_{max(w)+1} \vdash S_i \to \neg A^\sigma(w) & \textit{if } i \nvDash A \end{cases}$$

Proof. We prove the lemma by induction on A. If A is an atom and $i \vDash A$, then by the definition we have $T_0 \vdash S_i \to A^\sigma$. If $i \nvDash A$ then all j's in $A^\sigma = \bigvee_{j \vDash A} S_j$ are different from i, and by (i) in lemma 5.4, we conclude $T_0 \vdash S_i \to \neg A^\sigma$. The proof for the boolean cases is easy. For the modal case, if $i \vDash \Box B$, then for all j which $(i, j) \in R$, we have $j \vDash B$. Since w is a witness for $\Box B$, it is equal to (n, u) where n is greater than all the numbers in u. Therefore by IH, $T_{max(u)+1} \vdash S_j \to B^\sigma(u)$ for all j above i. Hence,

$$T_{max(u)+1} \vdash \bigvee_{(i,j) \in R} S_j \to B^\sigma(u).$$

Since $n \geq max(u) + 1$, we have

$$T_n \vdash \bigvee_{(i,j) \in R} S_j \to B^\sigma(u).$$

Then

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigvee_{(i,j) \in R} S_j \to B^\sigma(u)),$$

and consequently,

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bigvee_{(i,j) \in R} S_j) \to \mathrm{Pr}_n(B^\sigma(u)).$$

By (ii) in lemma 5.4, we have

$$T_{n+1} \vdash S_i \to \mathrm{Pr}_n(B^\sigma(u)),$$

and $n = max(w)$. Thus, the proof for this case is finished.

Now if $i \nvDash \Box B$, then there exists j which $(i,j) \in R$ and $j \nvDash B$. Again we have $w = (n, u)$, such that n is greater than all the numbers in u. Now by IH, $T_{max(u)+1} \vdash S_j \to \neg B^\sigma(u)$. Since $n \geq max(u) + 1$,

$$T_n \vdash S_j \to \neg B^\sigma(u)$$

and

$$I\Sigma_1 \vdash \mathrm{Pr}_n(B^\sigma(u) \to \neg S_j)$$

and then

$$I\Sigma_1 \vdash \neg\mathrm{Pr}_n(\neg S_j) \to \neg\mathrm{Pr}_n(B^\sigma(u))$$

and by (iii) in lemma 5.4, we have

$$T_{n+1} \vdash S_i \to \neg\mathrm{Pr}_n(B^\sigma(u))$$

and again since $n = max(w)$, the proof is complete. $\qquad\square$

Now, we state and prove the completeness theorem.

Theorem 5.6. *(Completeness) Let $(\mathbb{N}, \{T_n\}_{n=0}^\infty)$ be a reflexive provability model. If $(\mathbb{N}, \{T_n\}_{n=0}^\infty) \vDash A$, then $\mathbf{S4} \vdash A$. Therefore, if $\mathbf{Ref} \vDash A$, we have $\mathbf{S4} \vdash A$.*

Proof. Since $(\mathbb{N}, \{T_n\}_{n=0}^\infty) \vDash A$, there are extensions $B_1, \ldots, B_k$ of A and witnesses $w_1, \ldots, w_k$ such that for all arithmetical substitutions like σ, we have $\mathbb{N} \vDash \bigvee_{i=0}^k B_i^\sigma(w_i)$. Define $C = \bigvee_{i=0}^k B_i$ and $w = (w_i)_{i=0}^k$. Therefore, we know that w is a witness for C in $(\mathbb{N}, \{T_n\}_{n=0}^\infty)$. Now, we claim that $\mathbf{S4} \vdash C$. Pick N greater than all the numbers in w. If $\mathbf{S4} \nvdash C$ then there exists a finite reflexive transitive tree with clusters like (K, R, V), such that in one of the nodes in the root cluster (say k), C is false. Then by lemmas 5.4 and 5.5, we can construct an arithmetical substitution, such that $T_{max(w)+1} \vdash S_k \to \neg C^\sigma(w)$. Since the model is a reflexive provability model, all T_m's are sound and hence $\mathbb{N} \vDash S_k \to \neg C^\sigma(w)$. But by lemma 5.4 we know that $\mathbb{N} \vDash S_k$, thus $\mathbb{N} \vDash \neg C^\sigma(w)$, which contradicts with the assumption $\mathbb{N} \vDash C^\sigma(w)$. Therefore, $\mathbf{S4} \vdash C$. And finally, since in the presence of the axiom $\mathbf{K}$, all the extensions of a formula are equivalent to the formula itself, we have $\mathbf{S4} \vdash A$. For the second part of the theorem, it is easy to verify that that if $\mathbf{Ref} \vDash A$, then at least for one of the provability models like $(\mathbb{N}, \{T_n\}_{n=0}^\infty)$ we have $(\mathbb{N}, \{T_n\}_{n=0}^\infty) \vDash A$. And then the claim follows from the first part. $\qquad\square$

With the same techniques as above, we can strengthen the completeness theorem to the uniform and strong uniform completeness theorems. But first, we need a uniform version of lemma 5.3.

Definition 5.7. A hierarchy $\{T_n\}_{n=0}^{\infty}$ of theories is called uniform if there exists a Σ_1 formula $\mathrm{Prf}(x, y, z)$ such that for any n, m and A, $\mathrm{Prf}(n, m, \lceil A \rceil)$ iff m is a code of a proof for A in T_n. The hierarchy is called uniformly increasing if it is a uniform hierarchy and also we have $I\Sigma_1 \subseteq T_0$ provably in $I\Sigma_1$ and $I\Sigma_1 \vdash \forall x \forall z (\exists y\, \mathrm{Prf}(x, y, z) \to \exists w\, \mathrm{Prf}(x + 1, w, z))$ and finally it is called uniformly reflexive hierarchy if it is a uniformly increasing hierarchy such that for any formula A, $I\Sigma_1 \vdash \forall x \exists y\, \mathrm{Prf}(x+1, y, \exists w\, \mathrm{Prf}(x, w, A) \to A)$.

Lemma 5.8. *Let $\{T_n\}_{n=0}^{\infty}$ be a uniformly reflexive hierarchy of theories. Then, there is an arithmetical sentence $A(x, y)$ such that:*

   (i) $I\Sigma_1 \vdash \forall x, z \leq y\ (x \neq z \wedge A(x, y) \wedge A(z, y) \to \bot)$

   (ii) For all m, $I\Sigma_1 \vdash \bigvee_{i=1}^{m} A(i, m)$

   (iii) For any n, and any $i \leq m$, $T_{n+1} \vdash \neg\mathrm{Pr}_{T_n}(\neg A(i, m))$

   (iv) If we also assume that all theories in the hierarchy are consistent, then for any n, and any $i \leq m$, $\mathbb{N} \vDash \neg\mathrm{Pr}_{T_n}(\neg A(i, m))$ and $\mathbb{N} \vDash A(m, m)$.

Proof. The proof is basically the same as the proof of lemma 5.3. The only difference is that, here we have to define everything uniformly. First of all we need to define the hierarchy T'. Since T is a uniformly reflexive hierarchy, it is easy to prove that the hierarchy T' is a uniform hierarchy. Note that the definition of this new hierarchy is also uniform in p, i.e. there exists a proof predicate $\mathrm{Prf}(x, y, z, t)$ which means that y is a proof for z in T'_x when we choose t as our p. Now define, $B(x, y)$ as the following:

$$B(x, y) = \exists z \geq 1\ (\mathrm{Cons}(T'_{2zy-2x}) \wedge \neg\mathrm{Cons}(T'_{2zy-2x+1})),$$

and

$$A(x, y) = \forall 1 \leq z \leq x - 1\ \neg B(z, y) \wedge B(x, y).$$

Note that $A(x, y)$ and $B(x, y)$ are the uniform versions of A_r and B_r in which x stands for the index r and y for the number m. Now the proof of the properties we claimed is exactly same as the proof of lemma 5.3. The reason is that all properties are based on the standard numbers n, i and m. The only exception is (i), which is easily proved from the definition. $\quad\square$

Theorem 5.9. *(Uniform Completeness) Let $\{T_n\}_{n=0}^{\infty}$ be a uniform reflexive hierarchy of sound theories. Then there exists an arithmetical substitution $*$, such that for any modal formula A, if there exists a witness w such that for all $M \vDash \bigcup_n T_n$, $(M, \{T_n\}_{n=0}^{\infty}) \vDash A^*(w)$ then $\mathbf{S4} \vdash A$.*

Proof. First, note that according to the filteration method (see [4]), there exists a primitive recursive algorithm which reads A as an input and constructs a counter model (finite transitive reflexive tree with clusters) for A if $\mathbf{S4} \nvdash A$, and outputs zero, otherwise. Call this primitive recursive function, f. Therefore, if we use A_a to emphasize that the code for A is a, we have $f(a) = (W_a, R_a, V_a, w_a)$ in which w_a is a node in the root cluster such that $w_a \nvDash A_a$. The reason why such an f exists is that the size of a counter model is elementary bounded by the size of the code of the formula. (See [4].) Now assume that the function $\langle \cdot, \cdot \rangle$ is some canonical pairing function which is primitive recursive. Define $g(a)$ as the following primitive recursive function: Compute $f(a)$, change the name of all nodes like w in W_a to $\langle w, a \rangle$ and code the whole model again.

Now pick all $g(a)$'s and put all of them over one new reflexive root, k; and for valuation, use the induced valuation of the model plus the fact that the node k does not accept any atom. Then, use the technique of lemma 5.4 and define the function h on the whole new model:

$$h(0) = k \text{ and } h(x+1) = \begin{cases} j & \text{if } R(h(x), z) \quad \text{and} \ \ \mathrm{Prf}_T(x, \neg S(z)) \\ h(x) & \text{otherwise} \end{cases}$$

Where firstly, $T = \bigcup_{n=0}^{\infty} T_n$. It is easy to check that since the hierarchy is uniform, its union is also a recursively enumerable theory which has the following property: $I\Sigma_1 \vdash \mathrm{Pr}_n(A) \to \mathrm{Pr}_T(A)$. Secondly, $R(y, z)$ is a primitive recursive relation (Δ_1 formula in $I\Sigma_1$) which reads nodes y and z and if $y \neq k$, it decides whether they belong to the same model $g(pr_0(z))$, and if yes, whether (y, z) belongs to the relation of $g(pr_0(z))$, i.e. $R_{g(pr_0(z))}$. And if $y = k$, then the relation $R(y, z)$ decides whether z is in the $g(pr_0(z))$ or not (where $pr_0(z)$ is the index of the model which z belongs to). This R is a formalization of the accessibility relation of the new model. Note that we have to choose R in a way that the following holds:

(i) $I\Sigma_1 \vdash \forall x, y, z \ (R(x, y) \wedge R(y, z) \to R(x, z))$

(ii) For any node $i \neq k$, $I\Sigma_1 \vdash \forall x (R(i, x) \to \bigvee_{R_{g(pr_0(i))}(i,j)} x = j)$

It is easy to find such an R. The idea is, first using g to define a primitive recursive function $H(z)$ which reads z and outputs the whole set above z. Then define $R(x, y)$ as the existence of a sequence w from x to y such that for any r, w_{r+1} belongs to $H(w_r)$. The proof for these two properties are starightforward. (i) holds because of our transitive definition of R. (ii) needs the claim that if w is a sequence from i to x, then $x \in H(i)$. Use induction on the length of w to prove the claim.

And finally, the formula

$$S(z) = \exists y \forall x \geq y h(x) \in I(z) \wedge A(z, Card(I(z))) \wedge z = z$$

where $I(z)$ is a primitive recursive function, which reads z and computes the whole cluster of z. Note that here we use a uniform version of S_i's, and consequently we need the uniform version of A_r's. Now, for any $i \neq k$, the model above w_i is a finite reflexive transitive tree with clusters, and hence with the same arguments, we have the following:

(i) $T_0 \vdash \forall x, y \ (x \neq y \rightarrow (S(x) \rightarrow \neg S(y)))$.

(ii) $T_{n+1} \vdash S(i) \rightarrow \mathrm{Pr}_n(\bigvee_{(i,j) \in R} S(j))$ for all $i \neq k$.

(iii) If $(i, j) \in R$ then $T_{n+1} \vdash S_i \rightarrow \neg \mathrm{Pr}_n(\neg S_j)$ for all i.

(iv) $\mathbb{N} \vDash S_k$.

Since the model above any node $i \neq k$ is a finite model, the proof is the same as the proof of lemma 5.4, with only some minor changes. Firstly, for (i), we need the uniform version of the proof of lemma 5.4. It is implied by the facts that h is a provably total function in $I\Sigma_1$ and also the part (i) in lemma 5.8. Secondly, for (ii), we need to prove that if the function reaches i, then the limit cluster exists and it is above the cluster $I(i)$. It should be provable in $I\Sigma_1$. The idea is based on the fact that h is increasing and also the fact that if h reaches i, we can find the elements above i. These simple facts are provable by two properties of R which are mentioned before.

Now define the arithmetical interpretation as follows: $p^* = \exists z \ S(z) \wedge V(z, p)$ where $V(z, p)$ is a primitive recursive predicate (i.e. a Δ_1 formula in $I\Sigma_1$) which reads z and p and if $z \neq k$ decides whether p is true in the node z in the model $g(a)$, where $a = pr_0(z)$ is the index of the model which z belongs to. And if $z = k$, then rejects for all p. Since g is primitive recursive, this primitive recursive predicate exists. Note that V is a formalization of the valuation of the new model.

Now by a similar proof of lemma 5.5 we know that for all $i \neq k$, we have

$$\begin{cases} T_{max(w)+1} \vdash S_i \rightarrow A^\sigma(w) & \text{if } i \vDash A \\ T_{max(w)+1} \vdash S_i \rightarrow \neg A^\sigma(w) & \text{if } i \nvDash A \end{cases}$$

Now if $\mathbf{S4} \nvdash A$, then $i = w_a \nvDash A$, where a is the code of A. We have

$$T_{max(w)+1} \vdash S_i \rightarrow \neg A^*(w).$$

41

Hence for all $n \geq max(w) + 1$,

$$T_n \vdash S_i \rightarrow \neg A^*(w).$$

Then by

$$T_{n+1} \vdash S_k \rightarrow \neg \mathrm{Pr}_n(\neg S_i),$$

we have

$$T_{n+1} \vdash S_k \rightarrow \neg \mathrm{Pr}_n(A^*(w)).$$

Now since T_{n+1} is sound, $\mathbb{N} \vDash \neg \mathrm{Pr}_n(A^*(w))$ which means $T_n \nvdash A^*(w)$, and since n could be any big number, $T \nvdash A^*(w)$, therefore, there is a model of $T = \bigcup_n T_n$ like M, such that $M \nvDash A^*(w)$, which is a contradiction. Hence, $\mathbf{S4} \vdash A$. $\qquad \square$

Now, we are able to prove the strong completeness theorem.

Theorem 5.10. *(Uniform Strong Completeness) Let $\{T_n\}_{n=0}^{\infty}$ be a uniformly reflexive hierarchy of sound theories. Then there exists an arithmetical substitution $*$, such that for any modal sequent $\Gamma \Rightarrow A$, if there exist witnesses u and v such that for all $M \vDash \bigcup_n T_n$, $(M, \{T_n\}_{n=0}^{\infty}) \vDash \Gamma^*(u) \Rightarrow A^*(v)$, then $\mathbf{S4} \vdash \Gamma \Rightarrow A$. Moreover, If $\mathbf{Ref} \vDash \Gamma \Rightarrow A$, then $\Gamma \vdash_{\mathbf{S4}} A$.*

Proof. Use the arithmetical substitution from the uniform completeness. Now, since

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \Gamma^*(u) \Rightarrow A^*(v)$$

for all $M \vDash \bigcup_n T_n$, then $\bigcup_n T_n + \Gamma^*(u) \vdash A^*(v)$. Therefore, there is a finite subset $\Delta \subseteq \Gamma$ and a witness w, which is a subset of u, such that $\bigcup_n T_n + \Delta^*(w) \vdash A^*(v)$. Thus, for all $M \vDash \bigcup_n T_n$, we have

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \Delta^*(u) \Rightarrow A^*(v).$$

By uniform completeness, we have $\mathbf{S4} \vdash \Delta \Rightarrow A$ and hence, $\mathbf{S4} \vdash \Gamma \Rightarrow A$. The second part of the theorem, is obvious from the first part; because if $\mathbf{Ref} \vDash \Gamma \Rightarrow A$, then the assumption of the first part is true for some sequence of extensions like $\bar{\Gamma}$ and $B_1, B_2, \ldots, B_r$. Hence $\bar{\Gamma} \vdash_{\mathbf{S4}} \bigvee_{i=0}^{r} B_i$. Now, since in the presence of the axiom $\mathbf{K}$, the extensions of a formula are equivalent to the formula itself, we have $\Gamma \vdash_{\mathbf{S4}} A$. $\qquad \square$

6 The Logics GL and GLS

As Solovay showed in his pioneering work, [10], the logic $\mathbf{GL}$ is sound and complete for the interpretation that interprets all boxes as provability predicates in some appropriate theory. Moreover, he showed that if we change the

definition slightly, we can also capture the logic **GLS**. We translate his results into our framework and after defining constant and sound-constant provability models, we will show the soundness and completeness of **GL** and **GLS** for the classes of all constant provability models and all sound-constant provability models, respectively. In fact, the soundness-completeness theorems of these logics are just a new representation of Solovay's results. Consequently, we can claim that our provability interpretation is actually a generalization of Solovay's provability interpretation.

Definition 6.1. A provability model, $(M, \{T_n\}_{n=0}^{\infty})$ is constant if for any n and m, $(M, \{T_n\}_{n=0}^{\infty})$ thinks that $T_n = T_m$, i.e. $M \vDash \Pr_{T_m}(A) \leftrightarrow \Pr_{T_n}(A)$ and $M \vDash \Pr_{T_0}(\Pr_{T_m}(A) \leftrightarrow \Pr_{T_n}(A))$; and it is called a sound-constant model when it is constant and for any n, M thinks that T_n is sound, i.e. $M \vDash \Pr_{T_n}(A) \rightarrow A$. The class of all constant provability models and the class of all sound-constant provability models will be shown by **Cst** and **sCst**, respectively.

Remark 6.2. In the previous definition we used a notion for the equality of theories which seems ad-hoc and artificial. Here in this remark, we will justify that definition. Intuitively, M thinks that two theories are equal, when their provability-based properties are the same. In a more precise way, we say that M thinks T_n and T_m are equal, when for any modal sentence $\phi(p)$, any witness w and any arithmetical substitution σ for all atoms except p, $M \vDash \phi^{\sigma}(\Pr_m(A))(w) \leftrightarrow \phi^{\sigma}(\Pr_n(A))(w)$. Now, we will show that this definition of equality is equivalent to the original one. First of all, if we use $\phi(p) = p$, we will have $M \vDash \Pr_{T_m}(A) \leftrightarrow \Pr_{T_n}(A)$. Moreover, if we use $\phi(p) = \Box(p \leftrightarrow q)$, $w = (0)$ and σ where $q^{\sigma} = \Pr_n(A)$, we have $M \vDash \Pr_{T_0}(\Pr_{T_m}(A) \leftrightarrow \Pr_{T_n}(A))$. For the converse, we use induction on A to show the following claim.

    Claim. For any formula $\phi(p)$, any witness w and any arithmetical substitution σ for all atoms except p, M thinks that both of the following statements are true: $\phi^{\sigma}(\Pr_m(A))(w) \leftrightarrow \phi^{\sigma}(\Pr_n(A))(w)$ and $T_0 \vdash \phi^{\sigma}(\Pr_m(A))(w) \leftrightarrow \phi^{\sigma}(\Pr_n(A))(w)$.

    The atomic case and the boolean case are obvious. For the modal case, it is an easy consequence of the fact that Σ_1-completeness and some basic facts about the provability predicate are true in M.

    Now we are ready to prove the soundness-completeness result for **GL**. First of all, a technical lemma.

Lemma 6.3. *Let $(M, \{T_n\}_{n=0}^{\infty})$ be a constant provability model. Then for any modal formula A, any witness w and any arithmetical substitution σ, if $\mathbf{0}$ assigns zero to all the boxes of A, then M thinks that both of the following statements are true: $A^{\sigma}(w) \leftrightarrow A^{\sigma}(\mathbf{0})$ and $T_0 \vdash A^{\sigma}(w) \leftrightarrow A^{\sigma}(\mathbf{0})$.*

Proof. Use induction on A. The case for the atoms and the boolean connectives are easy. For the modal case, if $A = \Box B$, and $w = (n, u)$, then by IH, M thinks $T_0 \vdash B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0})$. Hence $T_n \vdash B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0})$ and by Σ_1-completeness, $M \vDash \mathrm{Pr}_n(B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0}))$. Thus $\mathrm{Pr}_n(B^{\sigma}(u)) \leftrightarrow \mathrm{Pr}_n(B^{\sigma}(\mathbf{0}))$ is true in M. Now, since $\mathrm{Pr}_n(B^{\sigma}(\mathbf{0}))$ and $\mathrm{Pr}_0(B^{\sigma}(\mathbf{0}))$ are equivalent in M, we have

$$M \vDash \mathrm{Pr}_n(B^{\sigma}(u)) \leftrightarrow \mathrm{Pr}_0(B^{\sigma}(\mathbf{0})).$$

For the other part of the claim, for $\Box B$, we have $M \vDash \mathrm{Pr}_n(B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0}))$. Therefore by Σ_1-completeness, M thinks $T_0 \vdash \mathrm{Pr}_n(B^{\sigma}(u) \leftrightarrow B^{\sigma}(\mathbf{0}))$. Hence $T_0 \vdash \mathrm{Pr}_n(B^{\sigma}(u)) \leftrightarrow \mathrm{Pr}_n(B^{\sigma}(\mathbf{0}))$ is true in M. But we know that M thinks that

$$T_0 \vdash \mathrm{Pr}_n(B^{\sigma}(\mathbf{0})) \leftrightarrow \mathrm{Pr}_0(B^{\sigma}(\mathbf{0})),$$

therefore, M thinks that

$$T_0 \vdash \mathrm{Pr}_n(B^{\sigma}(u)) \leftrightarrow \mathrm{Pr}_0(B^{\sigma}(\mathbf{0})).$$

$\square$

Theorem 6.4. *(Soundness) If $\Gamma \vdash_{\mathbf{GL}} A$, then $\mathbf{Cst} \vDash \Gamma \Rightarrow A$.*

Proof. By the previous lemma, it is enough to prove the following statement: if $\Gamma \vdash_{\mathbf{GL}} A$, then for any arithmetical substitution σ, $M \vDash \Gamma^{\sigma}(\mathbf{0}) \Rightarrow A^{\sigma}(\mathbf{0})$. To simplify the proof, we use A^{σ} instead of $A^{\sigma}(\mathbf{0})$. Now, to prove the statement, we claim:

Claim. If $\mathbf{GL} \vdash A$, then for any arithmetical substitution σ, $I\Sigma_1 \vdash A^{\sigma}$.

We prove the claim by induction on the length of the proof. The case for the axioms is easy; the main point is Löb's axiom which is derived by Löb's theorem formalized in $I\Sigma_1$. For the case of the rules, the case for the modus ponens rule is easy; for the neccesitation rule, by IH we have $I\Sigma_1 \vdash A^{\sigma}$, then since $I\Sigma_1 \subseteq T_0$ we have $T_0 \vdash A^{\sigma}$ and hence by Σ_1-completeness, we have $I\Sigma_1 \vdash \mathrm{Pr}_0(A^{\sigma})$ which completes the proof.

Now for the soundness theorem, if $\Gamma \vdash_{\mathbf{GL}} A$ then there exists a finite subset of Γ like Δ such that $\mathbf{GL} \vdash \bigwedge \Delta \to A$. Then by the claim, we have $I\Sigma_1 \vdash \Delta^{\sigma} \to A^{\sigma}$. Thus for any model M, $M \vDash \Gamma^{\sigma} \Rightarrow A^{\sigma}$. $\square$

For the completeness of **GL** we have:

Theorem 6.5. *(Uniform Strong Completeness) Let $I\Sigma_1 \subseteq T$ be a r.e. Σ_1-sound theory and $\{T_n\}_{n=0}^{\infty}$ be a hierarchy of theories such that for any n, $T_n = T$, then there is an arithmetical substitution $*$ such that for any modal sequent $\Gamma \Rightarrow A$, if for all $M \vDash T$, we have $(M, \{T_n\}_{n=0}^{\infty}) \vDash \Gamma \Rightarrow A$, then $\Gamma \vdash_{\mathbf{GL}} A$. And especially, if $\mathbf{Cst} \vDash \Gamma \Rightarrow A$ then $\Gamma \vdash_{\mathbf{GL}} A$.*

Proof. Pick $*$ as the uniform arithmetical substitution in Solovay's completeness theorem for T (see [2]). Now pick $M \vDash T$, arbitrarily. We have $(M, \{T_n\}_{n=0}^{\infty}) \vDash \Gamma \Rightarrow A$, hence there are a sequence of extensions like $\bar{\Gamma}$ and extensions like $\{A_i\}_{i=0}^{r}$ of A and witnesses like u and w_i such that

$$M \vDash \bar{\Gamma}^*(u) \Rightarrow \bigvee_{i=0}^{r} A_i^*(w_i).$$

Since all the theories are equal, we can easily verify that for any formula B and any witness v, $B^*(v)$ is equivalent to B^*, where B^* means a combination of substituting all the atoms by $*$ and interpreting any box as the provability predicate for T. Then we have

$$M \vDash \bar{\Gamma}^* \Rightarrow \bigvee_{i=0}^{r} A_i^*.$$

Moreover, it is easy to prove that if B is an extension of C, then B^* is equivalent to C^* in $I\Sigma_1$ and hence $M \vDash \Gamma^* \Rightarrow A^*$. Since M is arbitrary, we have $T + \Gamma^* \vdash A^*$, therefore, there is a finite subsequence of Γ like Δ such that $T + \Delta^* \vdash A^*$. Then by Solovay's uniform completeness theorem, we have $\Delta \vdash_{\mathbf{GL}} A$, thus $\Gamma \vdash_{\mathbf{GL}} A$. For the second part of the theorem, it is easy to show that if $\mathbf{Cst} \vDash \Gamma \Rightarrow A$, then the assumption of the first part for $T = I\Sigma_1$ is true, and hence $\Gamma \vdash_{\mathbf{GL}} A$. $\square$

For the case of **GLS** we have:

Theorem 6.6. *(Soundness) If $\Gamma \vdash_{\mathbf{GLS}} A$, then $\mathbf{sCst} \vDash \Gamma \Rightarrow A$.*

Proof. If $\Gamma \vdash_{\mathbf{GLS}} A$, then there are formulas $B_1, B_2, \ldots, B_k$ such that $\Gamma \vdash_{\mathbf{GL}} \bigwedge_{i=1}^{k}(\Box B_i \rightarrow B_i) \rightarrow A$. By the proof of the soundness of **GL**, we know that for any constant provability model and any arithmetical substitution σ, $M \vDash \Gamma^\sigma + \bigwedge_{i=1}^{k}(\mathrm{Pr}_0(B_i^\sigma) \rightarrow B_i^\sigma) \Rightarrow A^\sigma$. Since $M \vDash \mathrm{Pr}_0(C) \rightarrow C$ for any C, we have $M \vDash \Gamma^\sigma \Rightarrow A^\sigma$. Now, use lemma 6.3 to change the index of the theories to an appropriate increasing sequence. $\square$

Moreover, we have the completeness theorem.

Theorem 6.7. *(Completeness) Let $I\Sigma_1 \subseteq T$ be a sound r.e. theory and $\{T_n\}_{n=0}^{\infty}$ be a hierarchy of theories such that for any n, $T_n = T$. If $(\mathbb{N}, \{T_n\}_{n=0}^{\infty}) \vDash A$, then $\mathbf{GLS} \vdash A$; and especially, if $\mathbf{sCst} \vDash A$, then $\mathbf{GLS} \vdash A$.*

Proof. By the assumption, we have $(\mathbb{N}, \{T_n\}_{n=0}^{\infty}) \vDash A$. Hence, there are extensions like $\{A_i\}_{i=0}^{r}$ of A and witnesses like w_i such that for all arithmetical substitutions like σ, $\mathbb{N} \vDash \bigvee_{i=0}^{r} A_i^{\sigma}(w_i)$. Since all the theories are equivalent, it is easy to show that for any formula B and any witness v, $B^{\sigma}(v)$ is equivalent to B^{σ}, where B^{σ} means a combination of substituting any atom by σ and interpreting any box as the provability predicate for T. Therefore, $\mathbb{N} \vDash \bigvee_{i=0}^{r} A_i^{\sigma}$. Moreover, it is easy to prove that if B is an extension of C, then B^{σ} is equivalent to C^{σ} in $I\Sigma_1$, hence $\mathbb{N} \vDash A^{\sigma}$. Now since σ is arbitrary, based on Solovay's second completeness theorem, $\mathbf{GLS} \vdash A$.

For the second part of the theorem, it is easy to verify that if $\mathbf{sCst} \vDash A$ then the assumption of the first part for $T = I\Sigma_1$ is true and hence $\mathbf{GLS} \vdash A$. $\quad\square$

7 The Extensions of KD45

Intuitively, the logic $\mathbf{S5}$ does not admit any provability interpretation. The informal reason is as follows: The axiom $\mathbf{5} : \neg \Box A \to \Box \neg \Box A$ simply states that if A is not provable in a theory T_n, then this fact will be provable in T_{n+1}, i.e.

$$T_n \nvdash A \Rightarrow T_{n+1} \vdash \neg \mathrm{Pr}_n(A).$$

Moreover, the axiom $\mathbf{T}$ asserts that all theories are sound, hence

$$T_n \nvdash A \Leftrightarrow T_{n+1} \vdash \neg \mathrm{Pr}_n(A).$$

Now, we can use the last equivalence and the fact that the theory T_{n+1} is recursively enumerable to find a decision procedure for the provability in the theory $I\Sigma_1 \subseteq T_n$, which is impossible.

The above argument is based on the axiom $\mathbf{5}$ and the fact that all theories are sound. But it is possible to weaken the soundness part to some kind of consistency assumption which generalizes the above argument to all extensions of the logic $\mathbf{KD45}$.

Theorem 7.1. *There is no provability model $(M, \{T_n\}_{n=0}^{\infty})$ such that*

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \mathbf{KD45}.$$

Hence, there are no provability models for any extension of the logic $\mathbf{KD45}$. Specially, $\mathbf{S5}$ does not have any provability interpretation.

Proof. The proof we present here is more complex than the natural proof of this theorem, because we use weaker assumptions than what is available in **KD45**. The reason of our interest in this more complex proof is that we will use the same proof for the case of the classical propositional logic, and in that case we just have access to these weaker assumptions.

We prove the claim by contradiction. Suppose that there is a provability model $(M, \{T_n\}_{n=0}^{\infty})$ such that $(M, \{T_n\}_{n=0}^{\infty}) \vDash$ **KD45**. First, we show that the following three statements are true in M, then we will use these statements to reach the contradiction.

(i) For any n, M thinks that $T_{n+1} \nvdash \mathrm{Pr}_n(\bot)$. (Weak version of the consistency assumption.)

(ii) For any n, there exist $N > n$ and $s < N$ such that M thinks that $T_N \vdash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot)) \to \mathrm{Pr}_s(\bot)$. (Weak version of the provability of the consistency assumption.)

(iii) There are m, n and k such that M thinks that for any arithmetical statement ϕ,

$$\neg\mathrm{Pr}_n(\phi) \to \mathrm{Pr}_{m+1}(\mathrm{Pr}_k(\phi) \to \mathrm{Pr}_m(\bot)).$$

(Weak version of the axiom **5**).

To prove (i), for any number n, define $\Box^n\top$ as follows: $\Box^0\top = \top$ and $\Box^{n+1}\top = \Box\Box^n\top$. Now, consider the formula $\neg\Box\Box(\bot \wedge \Box^n\top)$, which is a theorem of **KD45**. Therefore, we have extensions of this formula, of the form $\neg\Box \bigvee_{j=0}^{s_i} \Box \bigvee_{k=0}^{t_{ij}}(\bot \wedge B_{ijk})$ for $0 \le i \le r$, where B_{ijk} is an extension of $\Box^n\top$. Moreover, there are witnesses $w_i = (n_i, (m_{ij}, (u_{ijk})_{k=0}^{t_{ij}})_{j=0}^{s_i})$ for any of these extensions such that for any arithmetical substitution σ, we have

$$M \vDash \bigvee_{i=0}^{r} \neg\Box \bigvee_{j=0}^{s_i} \Box(\bigvee_{k=0}^{t_{ij}}(\bot \wedge B_{ijk}))^{\sigma}(w_i).$$

Since the number of the boxes in $\Box^n\top$ is n, and witnesses for these boxes should be increasing, we have $m_{ij} \ge n$ and hence $n_i \ge n + 1$. Now, define $M = min_{ij}(m_{ij})$ and $N = min_i(n_i)$. Since B_{ijk} is an extension of the theorem $\Box^n\top$, we can easily show that $B_{ijk}(u_{ijk})$ is provable in $I\Sigma_1$. Hence, it is easy to see that $M \vDash \neg\mathrm{Pr}_N(\mathrm{Pr}_M(\bot))$ and $N > M \ge n$. Therefore, if $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$, and since $N > M \ge n$, we have $\mathrm{Pr}_N(\mathrm{Pr}_M(\bot))$, which is a contradiction.

For (ii), apply the same method to the formula $\Box(\Box\Box(\Box\perp\wedge\Box^n\top)\to\Box\perp)$ which is again a theorem of **KD45**. Then there are extensions of the form $\Box\bigvee_{j=0}^{q_j}(\Box(\bigvee_{k=0}^{p_{ij}}\Box\bigvee_{l=0}^{t_{ijk}}B_{ijkl})\to\Box\perp)$ where B_{ijkl} is an extension of $\Box\perp\wedge\Box^n\top$ and there are witnesses $w_i=(n_i,(m_{ij},(r_{ijk},(u_{ijkl})_{l=0}^{t_{ijk}})_{k=0}^{p_{ij}},s_{ij})_{j=0}^{q_i})$ such that

$$M\vDash\bigvee_{i=0}^{r}(\Box(\bigvee_{j=0}^{q_j}(\Box(\bigvee_{k=0}^{p_{ij}}\Box\bigvee_{l=0}^{t_{ijk}}B_{ijkl})\to\Box\perp))^{\sigma}(w_i).$$

Once more, with the same reason as in the case (i), $n\le r_{ijk}<m_{ij}<n_i$. Now define $N=max_i(n_i)$, $r=min_{ijk}(r_{ijk})$, $m=min_{ij}(m_{ij})$ and $s=max_i(s_i)$. Hence $N>m,r,s$ and $m>r\ge n$. Since the theories in the hierarchy $\{T_n\}_{n=0}^{\infty}$ is provably increasing, it is easy to prove

$$M\vDash\mathrm{Pr}_N(\mathrm{Pr}_m(\mathrm{Pr}_r(\perp))\to\mathrm{Pr}_s(\perp)).$$

Because $m>r\ge n$, we have

$$M\vDash\mathrm{Pr}_N(\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\perp))\to\mathrm{Pr}_s(\perp)).$$

Since n is arbitrary, we have proved that for any n, there exists $N>n$, $s<N$ such that

$$M\vDash\mathrm{Pr}_N(\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\perp))\to\mathrm{Pr}_s(\perp)),$$

and this is what we wanted.

For (iii) we know that $\neg\Box p\to\Box(\Box p\to\Box\perp)$ is provable in **KD45** and consequently it is true in the model. Therefore, there are some extensions of the formula like $\neg\Box p\to\Box\bigvee_{j=0}^{s_i}(\Box p\to\Box\perp)$, and some witnesses like $(n_i,m_i,(k_{ij},l_{ij})_{j=0}^{s_i})$ for them, such that for any arithmetical substitution σ,

$$M\vDash\bigvee_{i=0}^{r}(\neg\Box p\to\Box\bigvee_{j=0}^{s_i}(\Box p\to\Box\perp))^{\sigma}(n_i,m_i,(k_{ij},l_{ij})_{j=0}^{s_i}).$$

Now, define $n=max_i(n_i)$, $k=min_{ij}(k_{ij})$, $m=max_i(k_i)$ and $l=max_{ij}(l_{ij})$. It is easy to show that

$$M\vDash\neg\mathrm{Pr}_n(p^{\sigma})\to\mathrm{Pr}_m(\mathrm{Pr}_k(p^{\sigma})\to\mathrm{Pr}_l(\perp)).$$

It is easily verified that we can increase m and l; therefore, w.l.o.g. we can assume that $m=l+1$. Now send p to ϕ to prove the claim, and this completes the proof of the statement (iii).

For the proof of theorem 7.1, we want to use these three statements to reach a contradiction. First of all, to simplify the proof, use the following notation. For any a and b, define the theory $T_{b_a} = T_b + \text{Cons}(T_a)$. Thus, by $\text{Pr}_{b_a}(A)$, we mean $\text{Pr}_{T_{b_a}}$. Now, (iii) would be equivalent to

$$M \vDash \neg\text{Pr}_n(p^\sigma) \to \text{Pr}_{m_l}(\neg\text{Pr}_k(p^\sigma)).$$

Put $\phi = \text{Pr}_{m_l}(\bot)$; therefore,

$$M \vDash \neg\text{Pr}_n(\text{Pr}_{m_l}(\bot)) \to \text{Pr}_{m_l}(\neg\text{Pr}_k(\text{Pr}_{m_l}(\bot))).$$

On other hand by the formalized Σ_1-completeness, we have

$$I\Sigma_1 \vdash \neg\text{Pr}_k(\text{Pr}_{m_l}(\bot)) \to \neg\text{Pr}_{m_l}(\bot),$$

hence,

$$T_{m_l} \vdash \neg\text{Pr}_k(\text{Pr}_{m_l}(\bot)) \to \neg\text{Pr}_{m_l}(\bot).$$

Moreover, by Σ_1-completeness, we have

$$I\Sigma_1 \vdash \text{Pr}_{m_l}(\neg\text{Pr}_k(\text{Pr}_{m_l}(\bot)) \to \neg\text{Pr}_{m_l}(\bot)).$$

Therefore,

$$I\Sigma_1 \vdash \text{Pr}_{m_l}(\neg\text{Pr}_k(\text{Pr}_{m_l}(\bot))) \to \text{Pr}_{m_l}(\neg\text{Pr}_{m_l}(\bot)).$$

And since $M \vDash I\Sigma_1$, we have

$$M \vDash \neg\text{Pr}_n(\text{Pr}_{m_l}(\bot)) \to \text{Pr}_{m_l}(\neg\text{Pr}_{m_l}(\bot)).$$

Based on Gödel's second incompleteness theorem formalized in $I\Sigma_1$, we can conclude

$$I\Sigma_1 \vdash \neg\text{Pr}_{m_l}(\bot) \to \neg\text{Pr}_{m_l}(\neg\text{Pr}_{m_l}(\bot)).$$

However, by (i), we have

$$M \vDash \neg\text{Pr}_{l+1}(\text{Pr}_l(\bot)),$$

hence $M \vDash \neg\text{Pr}_{m_l}(\bot)$. Now, since $M \vDash I\Sigma_1$,

$$M \vDash \neg\text{Pr}_{m_l}(\neg\text{Pr}_{m_l}(\bot)).$$

Therefore,

$$M \vDash \text{Pr}_n(\text{Pr}_{m_l}(\bot)),$$

and thus by definition of T_{m_l} we have

$$M \vDash \text{Pr}_n(\text{Pr}_m(\text{Pr}_l(\bot))).$$

Now by (ii), there is some $N \geq l$ such that $M \vDash \text{Pr}_N(\text{Pr}_{l+1}(\text{Pr}_l(\bot)) \to \text{Pr}_s(\bot))$. W.l.o.g. pick this $N \geq n$. Since $N \geq n$, $M \vDash \text{Pr}_N(\text{Pr}_m(\text{Pr}_l(\bot)))$, and therefore, $M \vDash \text{Pr}_N(\text{Pr}_s(\bot))$. Because $N > s$, we have $M \vDash \text{Pr}_N(\text{Pr}_{N-1}(\bot))$, which contradicts with (i), and the proof follows. $\qquad\square$

8 A Remark on the Logic of Proofs

As we mentioned in the introduction, and as far as we know, the only successful attempt to find a natural provability interpretation for **S4** and hence, a formalization for the BHK interpretation is done by Artemov [1] and is called the logic of proofs. In this section, we will look into this approach and investigate some of its advantages and disadvantages.

The main idea of the logic of proofs, **LP**, is using explicit proofs to avoid the non-standard proofs and hence to eliminate the incompleteness phenomenon. Now, let us give a more detailed account of this result. The language of **LP** is two sorted; one sort is for the explicit proofs and the other for the propositions. The first sort consists of proof terms constructed by the proof variables, proof constants and the proof connectives $+$, $\cdot$ and !, while the second sort contains terms constructed by the propositional variables, propositional connectives and the predicate $t : A$ in which t is a proof term and A is a proposition. Now, let us explain the intuitive meaning of these operations:

First of all we have to emphasize that in this interpretation, despite the usual case in mathematics, proofs can be multi-conclusion. To find a natural candidate for these multi-conclusion proofs, it is enough to consider any usual proof as a proof for all intermediate statements it uses to prove the conclusion. For instance, the usual proof $A_1, A_2, \ldots, A_n$ of A_n will be interpreted as a proof for all A_i's.

1. The operation "!". If t is a proof for A, then $!t$ is a proof for the fact that "t is a proof for A". Therefore, the operator ! is the proof checker and could be interpreted as a self-awareness operator.
2. The operation "$\cdot$". If t is a proof for $A \to B$, and s is a proof for A, then $t \cdot s$ is a proof for B. Intuitively, $\cdot$ means the application of Modus Ponens on the proofs.
3. The operation "$+$". $t + s$ means the union of the proofs t and s. Recall that our proofs are multi-conclusion and $t + s$ can be served as a proof for all conclusions of t and s. Therefore if t is a proof for A and s is a proof for B, then $t + s$ is a proof for both A and B. To gain a better understanding, if we use the canonical way of changing usual proofs to multi-conclusion proofs, i.e. reading a usual proof as a proof for all intermediate statements in the proof, then $t + s$ just means putting t and s together. This is exactly what the symbol $+$ suggests.
4. The predicate " : ". The intuitive meaning of $t : A$ is that t is a proof for A.

The formal system **LP** is a theory in this language to capture the intended meaning of the symbols defined above. The axioms are the following:

1. A finite complete set of axioms for the classical propositional logic for the language of **LP**,
2. $t : A \to A$,
3. $t : A \to B \to (s : A \to t \cdot s : B)$,
4. $t : A \to !t : t : A$,
5. $t : A \to s + t : A$,
6. $s : A \to s + t : A$.

The rules are the modus ponens and the neccesitation rule. The latter means that for any axiom A, we have $\vdash c_A : A$, where c_A is an appropriate constant exclusively used for A.

Now, the natural interpretation for **LP** would be based on the usual proofs in Peano arithmetic. To formalize this idea, first of all we need a proof predicate: A proof predicate is a provably Δ_1 formula (in **PA**) $\mathrm{Prf}(x, y)$ with some natural basic properties (which we skip here. See [1]), and the following fundamental property:

$$\mathbf{PA} \vdash A \iff \exists x \mathrm{Prf}(x, \lceil A \rceil).$$

Now, we want to interpret the language of **LP** with this natural provability interpretation. Define an arithmetical interpretation $*$ as the following: Firstly, it interprets $\cdot$, !, $+$ and constants as the recursive functions on proofs in **PA** in the intended way. For instance, the function for $\cdot$ i.e., $\cdot^*$, will be the recursive function which reads the codes of the proofs for A and $A \to B$ and replies the code of a proof for B. Why can we define such recursive functions? To show the fact that these functions exist, we need a proof; but here we just want to explain the main idea instead of a formal proof. For this reason, let us limit ourselves to the canonical proof predicate of **PA**. In this case, it can be easily shown that we can define these functions in a recursive way. For instance, if x and y are proofs for $A \to B$ and A respectively, for $\cdot^*(x, y)$ it is enough to put y after x and add the formula B at the end. This is obviously a proof for B and this process is clearly a recursive function. Moreover, note that for any c_A, c_A^* is one of the proofs for the axiom A^*. The existence of such a c_A^* also needs a proof, which we skip here. (See [1].)

Up to this point, we have interpreted all the proof connectives as recursive functions. Now, use these interpretations to interpret all proof terms t.

Note that for interpreting proof variables we use arbitrary natural numbers as the code of proofs. Now, extend the interpretation $*$ to formulas. The idea is just interpreting all atoms as arithmetical sentences, reading $t : A$ as the proof predicate $\mathrm{Prf}(t^*, \lceil A^* \rceil)$ and commute $*$ with all boolean connectives. For instance, the interpretation of $!x : p \to p$ would be $\mathrm{Prf}(!^*(n), \lceil \phi \rceil) \to \phi$ where the interpretations of x and p are n and ϕ, respectively.

These arithmetical interpretations are the natural and the concrete interpretations of the proofs, and in [1] Artemov proved that **LP** is sound and complete with respect to the class of these arithmetical interpretations.

Theorem 8.1. LP $\vdash A$ *iff* A^* *is true for all arithmetical interpretations* $*$.

So far, we have found a natural proof interpretation for the system **LP**. Now, finding a natural interpretation for **S4** into **LP** would be the next step. Subsequently, we can use the composition of these interpretations to find a proof interpretation for **S4** and hence for **IPC**. We do not go into detail about the interpretation of the modal language into the system **LP**, but the basic idea is the following: Interpret any box as the existence of a proof; thus, any modal sentence will be equivalent to a first order formula in the language of **LP**. Therefore, we have quantifiers everywhere and specially in the scope of the predicate ":". We know that there is no way to exchange the quantifiers with the proof predicate (which is the reason why the incompleteness phenomenon and non-standard proofs appear), but since we require all the codes of the proofs to be standard numbers, we extract all the quantifiers and convert the translated formula into the prenex form. Now, use the Skolemization technique to witness the existential quantifiers by the universal ones. These witnesses are called realizations. (This is where we essentially need "+". It is important to note that by using Skolemization, we usually find a finite set of different witnesses and then we can roughly use + to merge these finite witnesses into one.) Note that this is not how Artemov argues in [1]; however, we explained the realizations in the way that we think is more accessible and to show that why it is natural to have such a concept in the heart of the interpretation of the modal sentences. Now, let us illuminate the above interpretation by an example.

Example 8.2. Consider the modal formula $\neg \Box p \to \Box \neg \Box p$. First, we have to interpret all of the boxes as the existence of the proofs. Hence, we have

$$\neg \exists x : p \to \exists y : (\neg \exists z : p).$$

Then, by extracting the quantifiers, we have

$$\forall x \neg x : p \to \exists y \forall z \; y : \neg z : p,$$

52

which is equivalent to

$$\exists y \exists x \forall z (\neg x : p \rightarrow y : \neg z : p)).$$

And finally by witnessing y and x by some terms $t(z)$ and $s(z)$, we have

$$\neg s(z) : p \rightarrow t(z) : \neg z : p.$$

Now, this new formula is *a* realization for the modal formula $\neg \Box p \rightarrow \Box \neg \Box p$. Note that this realization is just *one* possible realization of the formula and if we change the witnessing terms $t(z)$ and $s(z)$, we can find different realizations for the same formula.

After introducing the realizations, Artemov proved the following:

Theorem 8.3. ([**1**]) **S4** $\vdash$ *A iff there exists some realization r such that* **LP** $\vdash A^r$.

In sum, we can say that Artemov used two ingredients to find a provability interpretation for **S4**. The first one is the interpretation of modal sentences via realizations into the system **LP**. (Here the main idea is the interpretation of the boxes as the existence of the standard proofs.) And the second ingredient is the interpretation of the system **LP** via natural arithmetical proof interpretations. Therefore, the main idea of what Artemov did, is to use the system **LP** as a bridge to interpret **S4** via arithmetical proof interpretations.

Now, let us explain the advantages of this approach. First of all, it uses the explicit proofs and by the method of using realizations, it makes sure that everything is a standard proof in this context. Therefore, this approach actually kills the effect of Gödel's incompleteness theorems and makes the proof interpretation more intuitive. Note that naturally, we do not count infinite non-standard proofs as proofs. Moreover, regardless of the relation between modal logics and explicit proofs, the system **LP** has its own applications. In fact, since it is a formal system for explicit proofs, it can be used as a theory to investigate the concept of proof and its natural calculus. Consequently, these formal systems are appropriate to investigate the formal verification in computer science or the behavior of justifications in formal epistemology .

However, this utopia of explicit proofs comes at a price. The price is a combination of two unintended properties: The first one is related to the fundamental change in the interpretation of the concept of provability and the second one is about the role of **LP** as an unbiased bridge. The problem is

that the bridge is not neutral and somehow reflects its own behavior, which is not what we wanted.

Let us explain the first property by a simple example: Consider the modal sentence $\Box\neg\Box p$. The intended meaning of this sentence is the existence of a proof that shows p is not provable. In other words, it states that there *exists* a proof which shows that for *any* possible proof for p, like x, x is not a proof for p. Now, let us use the logic of proofs' interpretation of the sentence. Since the occurrences of the inner and the outer box are negative and positive respectively, the meaning of the sentence is the existence of a term $t(x)$ such that $t(x) : \neg x : p$. Forgetting the condition that the term $t(x)$ should be a term in the language, it means that for all x, there exists a proof $y = t(x)$ which proves $\neg x : p$. In other words, it says that for *any* possible proof for p like x, there *exists* a proof which shows that x is not a proof for p. It is easy to check that while the first interpretation is an $\exists\forall$ statement, the second one is a $\forall\exists$ statement, and it is obviously weaker than the first one. In fact, when we claim that we have a proof for unprovability of p, we mean a fixed uniform proof of the fact and we do not mean a machine (term) to transform a possible proof of x to a proof y that shows x is not a proof for p. What we showed above is just the difference for one statement. Nevertheless, the argument actually works for different kinds of sentences. The reason is simple: Logic of proofs needs to kill the presence of non-standard numbers. For this matter, it pushes out all the quantifiers. (It also changes the order of quantifiers to find a functional interpretation of proofs.) Since quantifiers do not commute with proof predicates, the sentence before pushing out the quantifiers is different from the sentence after that. The first sentence is the intended interpretation of provability and the latter is what the logic of proofs interprets as the meaning of provability. While this new interpretation is interesting and useful, it is not the intended interpretation of the informal provability and hence not *the* interpretation of **S4**.

In the following, we accept the functional interpretation of provability as what the logic of proofs proposed and we want to investigate the role of terms which we ignored in the previous argument. Now let us explain the second property by a thought experiment: Think about the situation that you have another connective in the language of **LP**, like "?" with the following intuitive meaning: If t is not a proof for A, then $?t$ is a proof of the proposition that "t is not a proof for A". Add the axiom

$$\neg t : A \rightarrow ?t : \neg t : A$$

to the system **LP** and call it **JS5**. Now, let us think about the natural arith-

metical interpretation of this connective and this new system. Since we used explicit standard proofs, we know that there exists a recursive function which reads t and if it is not a proof for A, finds a proof of this fact. The reason is as follows: We know that $\mathrm{Prf}(x, y)$ is provably Δ_1, hence if $\neg\mathrm{Prf}(t, \lceil A \rceil)$, we have

$$\mathbf{PA} \vdash \neg\mathrm{Prf}(t, \lceil A \rceil).$$

Therefore, by the definition of a proof predicate we have

$$\exists x \mathrm{Prf}(x, \lceil \neg\mathrm{Prf}(t, \lceil A \rceil) \rceil).$$

Now, use an unbounded search to find this x. Since it exists, our program halts and finds it. Accordingly, we know that such a function exists. Now interpret ? as this recursive function. Thus, based on this new natural arithmetical interpretation, we can interpret the new axiom $\neg t : A \to ?t : \neg t : A$. Hence, we have a natural arithmetical interpretation for the system $\mathbf{JS5}$. On the other hand, one of the instances of the new axiom, i.e. $\neg z : A \to ?z : \neg z : A$ where z is a proof variable, is the realization of the modal axiom $\mathbf{5} : \neg\Box A \to \Box\neg\Box A$ in this new language. (Simply, put $t(z) = ?z$ and $s(z) = z$ in the example 8.2.) The above discussion means that we can find a very natural provability interpretation of the axiom $\mathbf{5}$. Recall that this axiom is not provable in $\mathbf{S4}$ and it seems contradictory with Artemov's completeness result. However, there is no contradiction. The reason is that "?" is not in the original language of $\mathbf{LP}$, and hence you can not use it as a witness in the realization.

This observation shows that the arithmetical interpretation actually interprets the axiom $\mathbf{5}$, but the lack of the appropriate symbol in $\mathbf{LP}$ interferes with this fact. Therefore, the system $\mathbf{LP}$ does not reflect the whole power of the explicit proofs; it just chooses the appropriate part to witness all the theorems of $\mathbf{S4}$ and nothing more than that. In other words, the formalization of the provability interpretation via the explicit proofs is very sensitive to the language you use. If you change the language, then with the same arithmetical interpretation, you will capture different modal logics. Therefore, we can conclude that the soundness-completeness result for $\mathbf{S4}$ with respect to this kind of arithmetical interpretations is a soundness-completeness result for the language you use and not the natural arithmetical interpretation you choose. Now, a natural question would be the following: If we eliminate this language barrier and make the relation between modal logics and arithmetical interpretations as "direct" as possible, then which modal logic corresponds to the whole power of the arithmetical interpretations of the proofs? By the direct connection, we roughly mean the following: For any modal sentence A, write it in the prenex form in a way that we defined before, and

then, instead of witnessing the existential quantifiers by some terms in some language, witness them by some *natural recursive functions* on the proofs in Peano arithmetic. Now define the logic E as the logic of all statements which are valid for these kind of arithmetical interpretations. Clearly, the question mentioned above is informal, but it is easy to verify that the answer is not **S4**. The reason is that we can find an appropriate way to interpret **5** as we have shown above. It is appropriate because there is no *a priori* reason to accept the recursive function ! and reject ?. The first one finds a proof for $\mathrm{Prf}(m, n)$ if $\mathrm{Prf}(m, n)$ is true and the second finds a proof for $\neg\mathrm{Prf}(m, n)$ if $\mathrm{Prf}(m, n)$ is false. Both of them are recursive and hence accessible for us as human beings. Note that Prf is a provably recursive predicate and finding a proof for $\mathrm{Prf}(m, n)$ or its negation are similar as the computational task. (In the modal setting, the axioms **4** and **5** are intuitively different because we read $\Box A$ as $\exists x \mathrm{Prf}(x, A)$ which makes the sentences Σ_1 and now the formula and its negation are different.)

In fact, we have a realization theorem which extends the connection between **S4** and **LP** to **S5** and **JS5**, respectively. In a more precise way, if **S5** $\vdash A$, then there is a realization r such that **JS5** $\vdash A^r$ (see [5]), and since we can interpret **JS5** in arithmetic, we are able to claim that the intended modal logic E is at least **S5**. Therefore, this kind of approach based on the explicit proofs can not be considered as a formalization of the provability interpretation of **S4** and hence it can not be a formalization of the BHK interpretation of the intuitionistic logic. In fact, in the case of the propositional logics, this interpretation can capture the whole classical propositional logic (since the classical propositional logic is a propositional counterpart of the system **S5** via Gödel's translation). And obviously, this is not what we intended to have from a formalization of the BHK interpretation.

In sum, the explicit proofs approach first kills all the quantifiers and puts some explicit witnesses for them. Therefore, it inevitably eliminates the computability based difference between provability and unprovability (Σ_1 vs Π_1) and maps both predicates to the boolean combinations of the explicit proof predicate, Prf which belongs to the class Δ_1. As a consequence, the axioms **4** and **5** become similar and hence arithmetical interpretations can interpret **S5** in a very natural way. Finally, to avoid this fact, the logic of proofs uses the language of **LP** to regain the difference between **4** and **5** by choosing what we need for **S4** and ignore the other natural functions which in this case is the function ?. Therefore, our argument shows that the approach of explicit proofs does not distinguish **4** from **5** in a *natural* and *essential* way.

Now, let us compare what we do in this paper with the approach of the explicit proofs. First of all, we use the canonical meaning of provability instead of the logic of proofs' functional interpretation. Moreover, we do not use any language as a bridge. Therefore, our soundness-completeness results represent the provability behavior of our arithmetical interpretations in a direct way. Secondly, to capture different modal logics, we impose different natural conditions on our provability models, specially on the hierarchy of the theories. Therefore, we can claim that our approach can characterize different modal logics based on their different provability natures. Thirdly, our interpretation is based on the implicit proofs approach and hence it is a natural generalization of Solovay's work on **GL**. But since the Löb axiom is based on the incompleteness phenomenon, the explicit approach does not capture it and thus does not accept Solovay's provability interpretation as a special case. Hence, the explicit approach can not serve as the general framework for provability interpretations.

9 BHK Interpretations

Briefly, what we are going to do in this section, is to introduce a formalization for the BHK interpretation. Indeed, we will generalize this goal to make a framework to formalize different kinds of provability interpretations which includes the BHK interpretation as a special case. Note that the usual BHK interpretation is not the unique provability interpretation of the propositional language; in fact, there are many of them. Some of them, can be characterized as the variants of the original BHK interpretation, and some can't. The reason is that those provability interpretations do not satisfy the intended philosophical conditions which we want them to have, but they are still provability interpretations and they need an exact formalization if we want to use them. Let us illuminate the idea with two examples. The first one is a controversial variant of the BHK interpretation; it is obtained from the original BHK interpretation after relaxing the condition which says that there does not exist a proof for $\bot$. This interpretation informally corresponds to the minimal propositional logic, **MPC**. The second example of the provability interpretation is also obtained from the original BHK interpretation, but now we read $\bot$ as the inconsistency, instead of the provability of the inconsistency. More precisely, and using the notation of Gödel's translation, we have $\bot^g = \bot$, where g stands for this new translation (which is different from what we used in the introduction). This provability interpretation can not be characterized as a variant of the BHK interpretation because of some

philosophical reasons, which we do not get into here.

In this section, we try to justify the claim that our provability interpretation can prepare an appropriate framework to formalize these different provability interpretations of the propositional logics. To implement this idea, we need two steps. First, we have to interpret all the connectives as what the provability interpretation demands; this step is done by a translation like the Gödel's translation. The second step is interpreting the provability predicates (i.e. boxes in the modal translation) as the classical provability of the classical theories. For that reason, we need a hierarchy of theories to formalize the hierarchy of the intuitive provabilities in the definition of the provability interpretation and also a model to evaluate the truth value of our statements. This second step is done by the provability models.

What we discussed above is the general framework. Now, let us come back to the specific case, which is the original BHK interpretation. Is there a *right* formalization for this interpretation? As we will show later, for different kinds of provability models, we have different BHK interpretations and these interpretations could show inherently different provability behaviors. Consequently, there are different formalizations for the BHK interpretation, instead of just a canonical one. The reason is that the BHK interpretation just interprets propositional connectives in a discourse of provability, but it does not say anything about the internal structure of the concept of provability. For instance, it does not say anything related to the power of the meta-theories compared to the lower theories. Now, since the BHK interpretation is the intended semantics for the intuitionistic logic, we have to accept that there could be different intuitionistic logics in terms of different interpretations of the power of our model and our theories. All of them are equally intuitionistic if we have just the BHK interpretation as the criterion.

Now, the natural question is that what these intuitionistic logics are if we impose some natural conditions on the behavior of our model and our theories.
In the following, we will show that for some natural classes of the provability models like the class of all models or the class of all reflexive models, we can characterize some propositional logics like **BPC** and **IPC**, respectively. For instance, in the case of reflexive models, the result shows that if we use the BHK interpretation with the philosophical commitment which states that all of the theories, meta-theories, meta meta theories and so on are sound and also, any meta-theory is powerful enough to prove the soundness of the lower theories, then the logic of the formulas which are valid under this kind of

BHK interpretation, is the usual propositional intuitionistic logic. But, if we choose the minimal power, which does not assume any non-trivial condition on the hierarchy of the meta-theories, then the logic will change to **BPC**. However, what is important here is that all of these logics could be characterized as intuitionistic logics. This fact can explain the reason behind the disputes about finding the correct formalization of the intuitionistic logic. For instance, in [9], Ruitenburg argues that the *truly* intuitionistic logic is not **IPC** and he proposed **BPC** as the right one. Our approach here has a plural nature, and it tries to explain why with the same informal semantics (the BHK interpretation) there are different proposed logics.

And finally, we can discuss if there are any BHK interpretations for classical logic. The intuitive answer to this question is no. The reason is that when we interpret the axiom of the excluded middle, we should have the following condition: *p is provable or its unprovability is provable.* This means that the meta-theory should be powerful enough to prove the unprovability of all unprovable formulas. As we saw in the case of the logic **S5**, it contradicts with the natural condition that all the theories should be recursively enumerable. Therefore, intuitively speaking, we have to say that classical logic is beyond the scope of the BHK interpretation. In the following, we will prove this fact, in a precise way.

Definition 9.1. A provability interpretation for the propositional language is a translation from the propositional language to the language of modal logics.

To illuminate the definition 9.1, let us introduce three provability interpretations as examples.

Definition 9.2. The BHK interpretation b is the following translation:

(*i*) $p^b = \Box p$ and $\bot^b = \Box \bot$

(*ii*) $(A \wedge B)^b = A^b \wedge B^b$

(*iii*) $(A \vee B)^b = A^b \vee B^b$

(*iv*) $(A \to B)^b = \Box(A^b \to B^b)$

(*v*) $(\neg A)^b = \Box(A^b \to \Box\bot)$

Our translation is the same as the usual one, except for the case of $\bot$, which is translated to $\bot$ in the usual translation. (The negation of a formula A is considered as $A \to \bot$ and it inherits this change in the translation from $\bot$.

) The reason for slightly changing the definition of the translation is because the usual translation can not capture the intended intuition of the BHK interpretation. Actually, the intended intuitionistic meaning of $\perp$, like all the other atomic formulas, is its provability. Therefore, the natural interpretation of $\perp$ is $\square \perp$. On the other hand, we know that the BHK interpretation claims that there is not any proof of $\perp$, which means $\neg\square \perp$. Based on these two observations, we can justify the usual translation of $\perp$ as $\square \perp \wedge \neg\square \perp$, which is the same as $\perp$. Nevertheless, we have to emphasize that the condition of the unprovability of the inconsistency is not related to the meaning of the connectives, and hence it should not interfere in the BHK interpretation; it is actually a commitment we impose on the discourse of the provability. In our terms, the unprovability of the inconsistency asserts that the theories and meta-theories are consistent and it is obviously a property of the provability model and not a property of the connectives which we want to define. Hence, to formalize the original BHK interpretation, we need two ingredients; one is the b translation which is the formalization of the implicit BHK interpretation, and the second is the consistency condition on the provability models. The following definition formally states the second condition.

Definition 9.3. A provability model $(M, \{T_n\}_{n=0}^{\infty})$ is called a BHK model if for any n, $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\perp))$.

Remark 9.4. It seems that the natural consistency condition would be the consistency of all the theories. Yet, it is not enough. For instance, it is possible that all the theories in the hierarchy are consistent, but some meta-theory thinks that the lower theory is inconsistent, which contradicts with what an intuitionist assumes. For the intuitionist, the hierarchy of theories are just different layers of the story of the mind, and obviously these stories must be consistent in accordance with the BHK interpretation. However, this condition should be mentioned in the story itself. One way is assuming that any meta-theory actually proves the consistency of the lower theories. This is a natural condition, but it imposes a strong commitment on our theories. To keep the commitments as minimal as possible, we believe that the right condition to impose on the theories is the weaker condition which states that any meta-theory does not think that the lower theory is inconsistent. As we will see, this weaker condition widens the horizon of the BHK interpretation to capture the basic propositional logic on the one hand, and avoid artificial and degenerate models in which we could capture classical logic, on the other.

Based on the aforementioned considerations, when we talk about the formalization of the BHK interpretation, we always refer to the BHK models. Now, let us formalize what we will call the weak BHK interpretation.

60

Definition 9.5. Let q be a new atom which does not belong to the propositional language. The weak BHK interpretation, w, is the following translation:

(i) $p^w = \Box p$ and $\bot^w = \Box q$

(ii) $(A \wedge B)^w = A^w \wedge B^w$

(iii) $(A \vee B)^w = A^w \vee B^w$

(iv) $(A \to B)^w = \Box(A^w \to B^w)$

(v) $(\neg A)^w = \Box(A^w \to \Box q)$

The translation is based on the idea that in this variant of the BHK interpretation, we eliminate the consistency condition from the discourse of the provability. As a result, with this interpretation the intuitionist can not distinguish the inconsistency statement from any other statements. Therefore, in her viewpoint, $\bot$ is just a new atomic sentence which could be provable.

And finally, we will define Gödel's translation as a technical tool and also to show that there could be different provability models apart from the BHK interpretations.

Definition 9.6. Gödel's provability interpretation, g, is the following translation:

(i) $p^g = \Box p$ and $\bot^g = \bot$

(ii) $(A \wedge B)^g = A^g \wedge B^g$

(iii) $(A \vee B)^g = A^g \vee B^g$

(iv) $(A \to B)^g = \Box(A^g \to B^g)$

(v) $(\neg A)^g = \Box(\neg A^g)$

Now, it is time to define the satisfaction of a propositional formula in a provability model with respect to some provability interpretation i.

Definition 9.7. Let i be a provability interpretation. Then, by an extension of a propositional formula A, and a witness for A under the interpretation i, we mean an extension and a witness for A^i. And by $(M, \{T\}_{n=0}^{\infty}, i) \vDash A$ we mean $(M, \{T\}_{n=0}^{\infty}) \vDash A^i$. Moreover, if C is a class of provability models, by (C, i) we mean $\{(M, \{T\}_{n=0}^{\infty}, i) \mid (M, \{T\}_{n=0}^{\infty}) \in C\}$ and by $(C, i) \vDash A$ we mean $C \vDash A^i$.

The next step is establishing a soundness-completeness of the provability interpretations which we defined. But first, we need a technical lemma.

As a folklore, translations b and g are used interchangeably, and there is no problem while we use the modal logic **KD4** and beyond. However, for the case **K4**, they are not equivalent and in fact, the translation b establishes a soundness-completeness property between **K4** and **BPC** while the translation g is not complete for the mentioned two logics. (The g translation of the formula $(\top \to \bot) \to p$ is provable in **K4** while it is not provable in **BPC**.) In the following lemma, we prove the completeness of g between **K4** and **BPC** $+ (\top \to \bot) \Rightarrow \bot$ and then we will use it to prove the completeness of b between **KD4** and **BPC** $+ (\top \to \bot) \Rightarrow \bot$.

Lemma 9.8. *If $\Gamma^g \vdash_{\mathbf{K4}} A^g$, then $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot \vdash \Gamma \Rightarrow A$.*

Proof. To prove the lemma, we need the following claim:

Claim. If $\mathbf{GK4} \vdash \Gamma^g, P \Rightarrow \Delta^g, Q$, where P and Q are multi-sets of atoms and $P \cap Q = \emptyset$, then $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot \vdash \Gamma \Rightarrow \bigvee \Delta$.

To prove the claim, we use the cut-free sequent calculus for **K4** and **GK4**, and the proof will be based on induction on the length of the proof in **GK4**. The case for the axioms and structural rules are easy. For the case of the propositional rules, the possible last rules are the rules for conjunction and disjunction. The reason is that all formulas are in the form of g-translated formulas and atoms which are not in the form of implicational formulas. Since g commutes with conjunction and disjunction, these steps are easy to verify and we omit them here. For the case of the modal rule, assume that the sequent $\Box\Phi \Rightarrow \Box A$ is proved by $\Box\Phi, \Phi \Rightarrow A$. Then, since all of the formulas in $\Box\Phi \Rightarrow \Box A$ are boxed, every one of them should be non-atomic. Therefore, all of them are g translated formulas, and hence they are boxed atoms, or boxed implications. Consequently, there are two cases:

1. A is an atom. Call it p. If there is a p in Φ, we can prove the claim by axiom. And if Φ does not have p, we will read the proof backwards to find somewhere to apply IH. First of all, since all formulas in Φ are atomic or implicational, we can represent Φ by the following set: $P \cup \{A_i^g \to B_i^g\}_{i=0}^r$. Our goal is to show that $P, \{A_i^g \to B_i^g\}_{i=0}^r \Rightarrow \bot$ in $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot$. Secondly, all implications should be introduced by left implication, or by weakening. Therefore, if we continue reading the proof backwardly, it is impossible to have any modal rule before introducing all implications with those two rules. Now, in each branch of the tree find the the first point after which

all implications are introduced. Note that under these points the tree could have the conjunctive rule, the disjunctive rule, axioms and structural rules, but not the modal rule. It is easy to see that the formula in the mentioned nodes are of the form of the claim. The reason again is that conjunction and disjunction commute with the translation and moreover, the process can not introduce any new atoms. Then, by IH we can show that the translated parts in all of those nodes are provable in $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot$. Now, since all conjunction and disjunction rules and also the structural rules and axioms are valid in $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot$, we can reconstruct the proof in $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot$. However, the only problem is the left implicational rule which is not valid in $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot$. But note that in this case those rules just work with non-nested implications, and hence we can prove $P, \{A_i \to B_i\}_{i=0}^r \Rightarrow \top \to \bot$. Yet, we have the rule $(\top \to \bot) \Rightarrow \bot$, therefore, we have $P, \{A_i \to B_i\}_{i=0}^r \Rightarrow \bot$, which is what we wanted.

2. The case for the implicational A is similar. $\qquad\square$

Theorem 9.9. (i) $\mathbf{BPC} \vdash A$ iff $\mathbf{K4} \vdash A^b$

(ii) $\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot \vdash A$ iff $\mathbf{KD4} \vdash A^b$

(iii) $\mathbf{IPC} \vdash A$ iff $\mathbf{S4} \vdash A^b$

(iv) $\mathbf{FPL} \vdash A$ iff $\mathbf{GL} \vdash A^b$

(v) $\mathbf{MPC} \vdash A$ iff $\mathbf{S4} \vdash A^w$

Proof. The proof of the soundness part is easy and routine. For the completeness part, the case (iv) is proved by Visser in [12]. The same proof also works for (i). (iii) is a well-known result. (See [11].) For (ii), if $\mathbf{KD4} \vdash A^b$, then since $\mathbf{KD4} \vdash A^b \leftrightarrow A^g$, we have $\mathbf{KD4} \vdash A^g$. Therefore, $\mathbf{K4} \vdash \neg\Box \bot \wedge \Box\neg\Box \bot \to A^g$. Hence, $\mathbf{K4} \vdash \Box\neg\Box \bot \to \Box(\top \to\bot) \vee A^g$ and then

$$\mathbf{K4} \vdash ((\top \to \bot) \to\bot)^g \to ((\top \to\bot) \vee A)^g.$$

Thus,

$$\mathbf{K4} \vdash ((\top \to \bot) \to\bot)^g \Rightarrow ((\top \to\bot) \vee A)^g.$$

By lemma 9.8,

$$\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot \vdash ((\top \to \bot) \to\bot) \Rightarrow ((\top \to\bot) \vee A),$$

hence,

$$\mathbf{BPC} + (\top \to \bot) \Rightarrow \bot \vdash A.$$

For the case (v), we know that **MPC** and **S4** are sound and complete with respect to the class of reflexive transitive and value persistent Kripke models. However, in the case of **MPC**, the nodes can also satisfy $\bot$. Now, soundness is again easy. For the completeness part, if we have a counter **MPC**-Kripke model for A, we can construct a counter **S4**-model for A^w in the following way: Use the same Kripke model, with the same values, but assume that q is true in a node, if $\bot$ is true in that node. Then, it is easy to show that if the first model is a counter example for A, then the new one is a counter example for A^w. This construction proves the completeness part. $\square$

Now, we can use the soundness and completeness of these translations to transfer our results from the modal setting to the propositional one.

Definition 9.10. The class **BHK** is the class of all BHK models and the class **cBHK** is the class of all BHK models which are constant.

Theorem 9.11. (i) **BPC** $\vdash A$ *iff* $(\mathbf{BHK}, b) \vDash A$.

  (ii) **BPC** $+ (\top \Rightarrow \bot) \to \bot \vdash A$ *iff* $(\mathbf{Cons}, b) \vDash A$.

  (iii) **IPC** $\vdash A$ *iff* $(\mathbf{Ref}, b) \vDash A$.

  (iv) **FPL** $\vdash A$ *iff* $(\mathbf{cBHK}, b) \vDash A$.

  (v) *Let* $(M, \{T_n\}_{n=0}^{\infty})$ *be a provability model. Then* $(M, \{T_n\}_{n=0}^{\infty}, b) \vDash$ **CPC** *iff there exists n such that* $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$. *Therefore, there is not any BHK interpretation for classical logic.*

Proof. Based on theorem 9.9, the soundness parts are just easy consequences of the soundness results for the corresponding modal logics. For the completeness part for (i), if $(\mathbf{BHK}, b) \vDash A$, then there are extensions for A^w like B_i's and a witness for $\bigvee B_i$, such that for all arithmetical substitutions like σ, and all BHK models $(M, \{T_n\}_{n=0}^{\infty})$, we have $M \vDash (\bigvee_{i=0}^{r} B_i)^{\sigma}(w)$. Now, let Γ be a sequence of infinite copies of $\neg \Box\Box \bot$ and u a witness, which witnesses each of these formulas by $(n+1, n)$. We claim that for any provability model $(M, \{T_n\}_{n=0}^{\infty})$ and any arithmetical substitution σ, we have $M \vDash \Gamma^{\sigma}(u) \Rightarrow (\bigvee_{i=0}^{r} B_i)^{\sigma}(w)$. If $M \vDash \Gamma^{\sigma}(u)$, then for any n, we have $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$. Hence, $(M, \{T_n\}_{n=0}^{\infty})$ is a BHK model and therefore, $M \vDash (\bigvee_{i=0}^{r} B_i)^{\sigma}(w)$. Now, we know $\mathbf{PrM} \vDash \Gamma \Rightarrow A^b$; therefore, by strong completeness for **K4**, we have $\Gamma \vdash_{\mathbf{K4}} A^b$. Thus, $\mathbf{K4} \vdash \neg\Box\Box \bot \to A^b$ and then, $\mathbf{K4} \vdash ((\top \to \bot) \vee A)^b$. By theorem 9.9, $\mathbf{BPC} \vdash (\top \to \bot) \vee A$, and therefore by some easy argument $\mathbf{BPC} \vdash A$.

The case (iv) also needs an argument exactly like for the case (i). Moreover, since the consistent and reflexive models admit the consistency condition of

the BHK interpretation, the cases (ii) and (iii) are just a combination of theorem 9.9 and the completeness results for the corresponding theories.

For (v) we need some justification. First of all we want to show that if for any n, $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$, then $(M, \{T_n\}_{n=0}^{\infty})$ is not a model for **CPC**. We prove this claim by contradiction. Assume that for any n, $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$ and $(M, \{T_n\}_{n=0}^{\infty}, b) \vDash$ **CPC**. Now, we want to show that all three statements of the proof of theorem 7.1 is also true in our case. Firstly, (i) is true by assumption. Secondly, consider the formula $\Box^n\top$ which is a translation of the propositional classical theorem $\top^n$ with the definition $\top^0 = \top$ and $\top^{n+1} = \top \to \top^n$. Therefore, the formula $\Box(\Box\Box(\Box\bot \wedge \Box^n\top) \to \Box\bot)$ is the translation of the tautology $((\top \to (\top \to (\bot \wedge \top^n))) \to \bot)$. Thus,

$$(M, \{T_n\}_{n=0}^{\infty}) \vDash \Box(\Box\Box(\Box\bot \wedge \Box^n\top) \to \Box\bot).$$

Since we used this formula to show (ii), we can claim that we also have (ii) here. Thirdly, we know that $p \vee \neg p$ is a theorem of **CPC**. Hence, $(M, \{T_n\}_{n=0}^{\infty}) \vDash (p \vee \neg p)^b$, which means $(M, \{T_n\}_{n=0}^{\infty}) \vDash (\Box p \vee \Box(\Box p \to \Box\bot))$. Therefore, (iii) is also true in M. Thus, we have a contradiction and it proves the claim.

For the converse, assume that there is some n such that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$; we will show that $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash$ **CPC**. First of all, to simplify the proof, define the complexity of any box as the maximum depth of the nested boxes in front of that box. For instance, the complexity of the inner box in $\Box(\Box p \wedge q)$ is zero, and the complexity of the outer box is one. Now define the canonical witness starting from n, as follows: Witness any box by its complexity plus n. It is easy to show that this witness is an ordered one, because the witness for any outer box is bigger than the witness for the inner boxes. Define A^σ as the formula resulted by substituting all the atoms by σ and witnessing all the boxes by the canonical witness starting from n. Now, it is easy to verify that for any propositional formula $A \to B$, $M \vDash ((A \to B)^w)^\sigma$. To show this, firstly, note that the following claim holds: For any propositional formula B,

$$I\Sigma_1 \vdash \mathrm{Pr}_n(\bot) \to (B^w)^\sigma.$$

The proof of the claim is based on induction on B and easily follows. Now, assume that the complexity of the outmost box in $\Box(A^w \to B^w)$ is $k \geq n+1$. (Since witnesses begin with n and there is at least one box in A^w, k is at least $n + 1$.) By Σ_1-completeness we have

$$I\Sigma_1 \vdash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot) \to (B^w)^\sigma),$$

65

and hence,
$$I\Sigma_1 \vdash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot)) \rightarrow \mathrm{Pr}_k((B^w)^\sigma).$$

Then since $M \vDash I\Sigma_1$, then

$$M \vDash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot)) \rightarrow \mathrm{Pr}_k((B^w)^\sigma).$$

We know that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$ and $k \geq n+1$; hence $M \vDash \mathrm{Pr}_k(\mathrm{Pr}_n(\bot))$. Therefore,

$$M \vDash \mathrm{Pr}_k((B^w)^\sigma),$$

and thus,

$$M \vDash \mathrm{Pr}_k((A^w)^\sigma \rightarrow (B^w)^\sigma),$$

and the proof follows.

Now, it is easy to check that for any formula B, there exists another formula C such that C is in the CNF form, in which all the literals are implicational formulas, positive atoms and $\bot$ and classically equivalent to B. Note that the process of constructing this C just uses the classical rules for conjunction and disjunction. Now, since w and the canonical witness respect the conjunction and disjunction and the their basic rules, $(B^w)^\sigma$ and $(C^w)^\sigma$ are equivalent in M. Suppose that $\mathbf{CPC} \vdash B$; we want to show that $M \vDash (B^w)^\sigma$. It is enough to show that $M \vDash (C^w)^\sigma$. Considering that all the literals in C are implicational formulas, positive atoms and $\bot$, the literals of C^b are translations of implications, boxed atoms or $\Box \bot$. If $M \nvDash (C^w)^\sigma$, there must be some clause in which all the literals are false. Since the translations of the implications are true in M, there has to be a clause in C consisting of atoms and $\bot$. Therefore, C can not be a classical tautology and hence B will not be, as well. But $\mathbf{CPC} \vdash B$; a contradiction. Thus, $M \vDash (B^w)^\sigma$. So far, we have shown that if $\mathbf{CPC} \vdash B$, then $M \vDash (B^w)^\sigma$. Now, if we send q in the definition of $\bot^w = \Box q$, to $\bot$, then we have $M \vDash (B^b)^\sigma$, which proves the theorem. $\qquad\square$

There is another type of the BHK interpretation in which there is not any kind of assumption on the non-existence of a proof of the contradiction.

Theorem 9.12. (i) $\mathbf{MPC} \vdash A$ *iff* $(\mathbf{Ref}, w) \vDash A$.

(ii) *Let* $(M, \{T_n\}_{n=0}^\infty)$ *be a provability model. Then* $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{IPC}$ *iff* $(M, \{T_n\}_{n=0}^\infty, w) \vDash \mathbf{CPC}$ *iff there exists* n *such that* $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$.

Proof. For (i), use theorem 9.9 and the soundness-completeness results for **S4**. For (ii), if there exists n such that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$, then by the proof

of theorem 9.11 part (v), we know that $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash \textbf{CPC}$. Moreover, if $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash \textbf{CPC}$, then we can easily verify that we have $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash \textbf{IPC}$. Now, it remains to show that if $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash \textbf{IPC}$, then there exists n such that $M \vDash \mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$.

Assume that $(M, \{T_n\}_{n=0}^{\infty}, w) \vDash \textbf{IPC}$ and for any n, $M \vDash \neg\mathrm{Pr}_{n+1}(\mathrm{Pr}_n(\bot))$. We want to reach a contradiction. We know that $\textbf{IPC} \vdash \bot \rightarrow p$. Hence, $(M, \{T_n\}_{n=0}^{\infty}) \vDash (\bot \rightarrow p)^w$. Thus, $(M, \{T_n\}_{n=0}^{\infty}) \vDash \Box(\Box q \rightarrow \Box p)$. Consequently, there are extensions of the form, $\Box(\bigvee_{j=0}^{s_i}(\Box q \rightarrow \Box p))$ for $0 \leq i \leq r$ and witnesses $w_i = (n_i, (m_{ij}, k_{ij})_{j=0}^{s_i})$ such that for any arithmetical substitution σ,

$$M \vDash \bigvee_{i=0}^{r} \Box(\bigvee_{j=0}^{s_i}(\Box q \rightarrow \Box p))^{\sigma}(w_i).$$

Now, define $k = max_{ij}(k_{ij})$, $m = min_{ij}(m_{ij})$ and $n = max_i(n_i)$. It is easy to see that

$$M \vDash \mathrm{Pr}_n((\mathrm{Pr}_m(q^{\sigma}) \rightarrow \mathrm{Pr}_k(p^{\sigma}))).$$

And if we choose a substitution σ such that $q^{\sigma} = (0 = 0)$ and $p^{\sigma} = (0 = 1)$, then we have

$$M \vDash \mathrm{Pr}_n((\mathrm{Pr}_m(0 = 0) \rightarrow \mathrm{Pr}_k(0 = 1)),$$

and hence $M \vDash \mathrm{Pr}_n(\mathrm{Pr}_k(\bot))$. Thus, for some number $N > n, k$, we have $M \vDash \mathrm{Pr}_{N+1}(\mathrm{Pr}_N(\bot))$ which is a contradiction. $\qquad\square$

References

[1] S. Artemov. Explicit provability and constructive semantics. Bulletin of Symbolic Logic, 7(1):1-36, 2001.

[2] G. Boolos, The logic of provability, Cambridge University Press, 1993.

[3] S. Buss, The modal logic of pure provability, Notre Dame Journal of Formal Logic, vol. 31 (1990), no. 2, pp. 225-231.

[4] A. Chagrov and M. Zakharyaschev, Modal Logic, Oxford University Press, 1997.

[5] M. Fitting, The Realization Theorem for S5 A Simple, Constructive Proof, Published in Game, Norms and Reasons, Springer, 2011, pp 61-76.

[6] K. Gödel, Eine Interpretation des Intuitionistichen Aussagenkalküls, Ergebnisse Math Colloq. Vol. 4 (1933), pp. 39-40.

[7] S. Kripke, Semantical considerations on modal logic, Acta Philosophica Fennica, vol. 16 (1963), pp. 83-94.

[8] F. Poggiolesi, Gentzen Calculi for Modal Propositional Logic, Springer, 2010.

[9] W. Ruitenburg, Constructive logic and the paradoxes, Modern Logic 1, No. 4 (1991), pp. 271-301.

[10] R. Solovay, Provability interpretations of modal logic, Israel Journal of Mathematics, vol. 25 (1976), pp. 287-304.

[11] A. S. Troelstra, H. Schwichtenberg, Basic Proof Theory, Second Edition, Cambridge University Press, 2000.

[12] A. Visser, A propositional logic with explicit fixed points. Studia Logica 40 (1981), 155-175.