The Closed Fragment of IL is PSPACE Hard

Félix Bou[1,2]

Department of Probability, Logic and Statistics
University of Barcelona
Barcelona, Spain

Joost J. Joosten[1,3]

Department of Logic, History and Philosophy of Science
University of Barcelona
Barcelona, Spain

Abstract

In this paper we consider $\mathbf{IL}_0$, the closed fragment of the basic interpretability logic $\mathbf{IL}$. We show that we can translate $\mathbf{GL}_1$, the one variable fragment of Gödel-Löb's provabilty logic $\mathbf{GL}$, into $\mathbf{IL}_0$. Invoking a result on the PSPACE completeness of $\mathbf{GL}_1$ we obtain the PSPACE hardness of $\mathbf{IL}_0$.

Keywords: Interpretability Logic, computational complexity, closed fragment, PSPACE, normal forms.

1 Introduction

For a propositional logic L, the closed fragment –we write L_0– of that logic consists of those theorems of L that do not contain any propositional variables at all. For various logics, it is known that the closed fragment is a lot easier than the full logic itself. The simplicity of the closed fragment can be captured by the complexity class of a decision procedure of theoremhood. Moreover, in all cases where L_0 is known to be simpler than L in this sense, we have a set of *normal forms* for L_0 and a normal form theorem to the effect that each closed formula can be written in a unique way as a special combination of normal form formulas.

Perhaps the most canonical example of this phenomenon is classical propositional logic. Theoremhood in classical propositional logic is known to be co-NP

---

complete whereas the closed fragment is decidable in LOG-time. In this case, by definition, the only two formulas in normal form are $\top$ and $\bot$. For various modal logics the situation is similar but slightly different. For the provability logic **GL**, theoremhood is known to be a PSPACE complete problem (see [2, Theorem 18.29]), whereas provability of formulas in the closed fragment is known to be PTIME decidable (see [3, Theorem 9]). Moreover, the normal form theorem ([1, Chapter 7]) states that each formula in the closed fragment is provably equivalent to a Boolean combination of formulas of the form $\Box^n \bot$ with $n \in \omega$.

Interpretability logics arise as natural extensions of **GL**. The logic **GL** has only one modal operator $\Box A$ to capture that "A is provable in some basic theory T". Interpretability logics have an additional binary modality $A \rhd B$ to capture that "the theory $T + A$ interprets the theory $T + B$".

These interpretability logics are always defined as some core part **IL** as defined below, together with some additional principles. However, as soon as the additional principles prove some rather weak principle F, the technical details of which are irrelevant for the moment, then closed interpretability formulas can be expressed without the modality $\rhd$ and the normal forms are the same as those of **GL**: Boolean combination of formulas of the form $\Box^n \bot$ with $n \in \omega$ (see [5]). It is good to stress here that all interpretability logics with some interesting meta-mathematical content do contain the principle F. For logics below **ILF** and in particular for **IL** itself, it is not known if there exists a natural set of normal forms.

Not for all modal logics it is the case that the L_0 is simpler than L. In particular, it is known that the minimal modal logic **K** and its closed fragment $\mathbf{K}_0$ are both PSPACE complete (see [3, Corollary 4]). The same also happens for the modal logic **K4** of transitive frames (see also [3]).

We shall see in this paper that the logic **IL** is like these logics **K** and **K4** in that also the closed fragment of **IL** is PSPACE hard thereby settling an open question in [8] in the negative as to whether the closed fragment allows a nice characterization.

2   Interpretability logics

Interpretability logics have been primarily used to study in a formalized setting the notion of relativized interpretability which is captured by a binary modal operator $\rhd$. The phrase $p \rhd q$ is to be read as "(T together with the translation of p) interprets (T together with the translation of q)" for some base theory T. Different theories T prove different modal principles to hold. However, all theories that allow for coding of syntax and thus for formalizing the notion of interpretability do validate some core logic which is called **IL**.

2.1   The logic **IL**

We recall that **GL** is the normal modal logic with one modality $\Box$ whose non-logical axioms are instantiations of the following axiom schemes.

(i) $\Box(A \to B) \to (\Box A \to \Box B)$

(ii) $\Box(\Box A \to A) \to \Box A$

It is well known that **GL** proves the transitivity axiom, that is,

$\Box A \to \Box\Box A$.

The logic **IL** is formulated in a propositional modal logic with two modalities $\Box$ and $\rhd$. We shall use the following reading conventions. The strongest binding operators are $\neg$ and $\Box$ followed by $\vee$ and $\wedge$ which in turn bind stronger than $\rhd$. The weakest binding connectives are the implications $\leftrightarrow$ and $\to$. We shall write $\Diamond\varphi$ as shorthand for $\neg\Box\neg\varphi$.

Definition 2.1 The logic **IL** is a normal modal logic containing **GL** whose rules are Modus Ponens and Necessitation and whose axioms other than all propositional tautologies are the instances of the following axiom schemes.

$J1$ $\Box(A \to B) \to A \rhd B$

$J2$ $(A \rhd B) \wedge (B \rhd C) \to A \rhd C$

$J3$ $(A \rhd C) \wedge (B \rhd C) \to A \vee B \rhd C$

$J4$ $A \rhd B \to (\Diamond A \to \Diamond B)$

$J5$ $\Diamond A \rhd A$

It follows from $J1$ and $J4$ that $\Box$ is expressible in terms of $\rhd$ within **IL**:

$\mathbf{IL} \vdash \Box A \leftrightarrow \neg A \rhd \bot$.

The logic **ILF** is obtained by adding the axiom F to **IL**.

$\mathsf{F} \quad := \quad \Diamond A \to \neg(A \rhd \Diamond A)$

This principle can be seen as a natural generalization of Gödel's second incompleteness theorem. Gödel's second incompleteness theorem states that any recursive theory, whenever consistent, does not prove its own consistency. The principle F states that any recursive theory, whenever consistent, does not even *interpret* its own consistency.

2.2   Semantics for **IL**

The logic **IL** allows for natural Kripke semantics where the binary modality $\rhd$ is modeled by a ternary relation. Rather than working with a ternary relation, we tend to conceive the semantics for $\rhd$ as a collection of binary relations.

Definition 2.2 An **IL** model, also called Veltman-model, is a quadruple $\langle W, R, \{S_x : x \in W\}, \Vdash \rangle$ where W is a non-empty set of worlds, R is a binary relation on W that is transitive and conversely well-founded. For each $x \in W$, the binary relation S_x is transitive and reflexive such that moreover

(i) $yS_x z \to xRy \wedge xRz$;

(ii) $xRyRz \to yS_x z$.

The relation $\Vdash$ is a usual forcing relation that can be conceived as a map assigning to each propositional variable p a subset $v(p)$ of W of the worlds where p holds. We

write $x \Vdash p$ to indicate that $x \in v(p)$. The relation $\Vdash$ is extended to the set of all formulae by stipulating that

(i) $x \Vdash \Box A \iff \forall y \ (xRy \to y \Vdash A)$;

(ii) $x \Vdash A \rhd B \iff \forall y(xRy \land y \Vdash A \to \exists z(yS_x z \land z \Vdash B))$.

It is well-known that **IL** is sound and complete with respect to the class of all Veltman models (see [4]).

2.3 Fragments

We shall denote by $\mathbf{IL}_0$ the fragment of **IL** that consists of those modal formulae provable in **IL** that contain no propositional variables. Likewise, by $\mathbf{GL}_1$ we shall denote those formulas in the language of **GL** that contain only one variable and are theorems of **GL**.

3   Translating $\mathbf{GL}_1$ into $\mathbf{IL}_0$

Let p be the variable of $\mathbf{GL}_1$. We shall translate this variable to some formula in the closed fragment of **IL** that essentially uses the $\rhd$ modality. It is easy to see that such formulas exist. Examples are given in [8] (Section 5.4) and in [7]. The formula that we use here is equal to the one exposed in [7].

3.1 Some motivation for our translation

In this section we shall expose a translation that reduces theoremhood of $\mathbf{GL}_1$ to $\mathbf{IL}_0$ thereby establishing PSPACE hardness of the latter. The motivation for this translation is mainly semantical.

We will code the information as to whether p holds or not in a world x by making the formula $\top \rhd \Diamond \top$ true at x if and only if $x \Vdash p$. To this extent we can glue to each x two new worlds x_1 and x_2 with xRx_1Rx_2 and [4] $x_2 S_x x_1 \iff x \Vdash p$. For this, all the S_x relations in **IL** are sufficiently independent. This idea should motivate why we translate p to $\Diamond\Diamond\top \to \top \rhd \Diamond\top$.

Moreover, with this approach the points that we are interested in, that is, the original points, become easily definable by the formula $\Diamond\Diamond\top$. Thus, when quantifying over points that we are interested in, we should relativize to our old domain. This explains why we shall translate $\Box A$ to $\Box(\Diamond\Diamond\top \to A^\dagger)$ where $A^\dagger$ is the translation of A.

We shall see in Subsection 3.3 that we do not actually need to glue so many different new worlds to code all the behavior of the $x \vdash p$ for all x in the model. By transitivity it suffices to add some worlds only at the top of the model.

---

[4] We should add some more S_x relations too on the already existing part of the model. For the motivational part here, we just focus on the newly added worlds x_1 and x_2.

3.2   The translation

We consider the following translation † of formulas of $\mathbf{GL}_1$ into formulas of $\mathbf{IL}_0$:

  (i) $\perp^\dagger = \perp$

  (ii) $p^\dagger = \Diamond\Diamond\top \to (\top \triangleright \Diamond\top)$

  (iii) $(A \to B)^\dagger = A^\dagger \to B^\dagger$

  (iv) $(\Box A)^\dagger = \Box(\Diamond\Diamond\top \to A^\dagger)$

Lemma 3.1 *Let A be a formula of* **GL** *that only contains the propositional variable p. If* $\mathbf{GL} \vdash A$, *then* $\mathbf{IL} \vdash A^\dagger$.

Proof. So, suppose $\mathbf{IL} \nvdash A^\dagger$. Then, there is an **IL** model $\mathcal{M} = \langle W, R, \{S_x : x \in W\}, \Vdash \rangle$ and a world $w \in W$ such that $\mathcal{M}, w \nVdash A^\dagger$. Next, we consider the **GL** model $\mathcal{N} = \langle W', R', \Vdash' \rangle$ defined by:

  (i) $W' := \{w\} \cup \{x \in W : \mathcal{M}, x \Vdash \Diamond\Diamond\top\}$,

  (ii) $R' := R \cap (W' \times W')$,

  (iii) $x \Vdash' p$ iff $\mathcal{M}, x \Vdash p^\dagger$ (for every $x \in W'$).

We point out that the union defining W' may be a non-disjoint one. Using the definition of $\mathcal{N}$ it is straightforward to prove (by induction on the length of the formula) that for every formula B which only contains the propositional variable p,

$$\mathcal{N}, x \Vdash' B \text{ iff } \mathcal{M}, x \Vdash B^\dagger \quad \text{(for every } x \in W').$$

In particular, we get that $\mathcal{N}, w \nVdash' A$. Therefore, $\mathbf{GL} \nvdash A$.   □

Lemma 3.1 is the easier direction of what we shall see is an equivalence. In particular, the lemma allows for an easy proof-theoretic proof.

Proof. So, suppose $\mathbf{GL} \vdash A$. We know that **GL** has a cut-free proof π (see system $\mathcal{G}^1$ in [6]) which thus satisfies the sub-formula property. Consequently, each sequent in π contains at most the variable p. It is an easy check that proofs only containing p are preserved under †.   □

3.3   Construction on models

In this subsection we shall prove the converse to Lemma 3.1.

Lemma 3.2 *Let A be a formula of* **GL** *that only contains the propositional variable p. If* $\mathbf{IL} \vdash A^\dagger$, *then* $\mathbf{GL} \vdash A$.

Proof. By the completeness proofs of **IL** and **GL** we know that it is sufficient to show that

$$\forall A \, [\forall^{\mathbf{IL}\text{-model}}\mathcal{M} \; \mathcal{M} \models A^\dagger \;\; \Longrightarrow \;\; \forall^{\mathbf{GL}\text{-model}}\mathcal{N} \; \mathcal{N} \models A],$$

or equivalently

$$\forall A \, [\exists^{\mathbf{GL}\text{-model}}\mathcal{N} \; \exists n \in \mathcal{N} \; \mathcal{N}, n \Vdash A \;\; \Longrightarrow \;\; \exists^{\mathbf{IL}\text{-model}}\mathcal{M} \; \exists m \in \mathcal{M} \; \mathcal{M}, m \Vdash A^\dagger].$$

To this extent we shall exhibit a transformation on **GL** models that yields the desired **IL** model. Let $\mathcal{N} := \langle W, R, \Vdash \rangle$ be a **GL**-model for some **GL**-formula A at most containing the variable p. As **GL** is complete with respect to finite tree-like models we may indeed assume that $\mathcal{N}$ has is such a finite tree-like model. By E=$\{e_1, \ldots, e_l\}$ we denote the set of end-points in $\mathcal{N}$, which is of course finite. We consider two disjoint copies of E (which are also disjoint with W), namely $E^\flat = \{e_1^\flat, \ldots, e_l^\flat\}$ and $E^\natural = \{e_1^\natural, \ldots, e_l^\natural\}$. The idea is to 'glue' these additional points $e_i^\flat$ and $e_i^\natural$ as a little R chain of length two above [5] the end-points e_i so that each old point in the model will satisfy $\Diamond\Diamond\top$.

As R is conversely well-founded it is the case that each $x \in M$ is $R^=$-below some e_i (here $xR^=y$ is a shorthand for "$xRy \lor x = y$").

Now, we consider an **IL** model $\mathcal{M} = \langle W', R', \{S_x' : x \in W'\} \rangle$ satisfying:

(i) $W' := W \uplus E^\flat \uplus E^\natural$,

(ii) R' is the transitive closure of $R \cup \{(e, e^\flat) : e \in E\} \cup \{(e^\flat, e^\natural) : e \in E\}$,

(iii) $S_{e^\flat}' = \{(e^\natural, e^\natural)\}$ (for every $e^\flat \in E^\flat$).

(iv) $S_{e^\natural}' = \emptyset$ (for every $e^\natural \in E^\natural$).

(v) for every $x \in W$ it holds that

S_x' is the smallest transitive and reflexive relation on $\{y \mid xR'y\}$ that contains both S_x and R' restricted to $\{y \mid xR'y\}$ such that moreover for every $e \in E$,
$$(e^\natural, e^\flat) \in S_x' \quad \text{iff} \quad \mathcal{N}, x \Vdash p \text{ and } xR^=e.$$

It is very easy to check that there is a unique **IL** model satisfying these conditions. We notice that in the definition of W' we have used the symbol $\uplus$ to emphasize that these unions are indeed disjoint ones, and we have not introduced a valuation $\Vdash'$ in $\mathcal{M}$ because our purpose is only to evaluate closed formulas (like $A^\dagger$).

First of all we note that in $\mathcal{M}$ we can modally define the old points in $\mathcal{N}$ since $\{x \in W' : \mathcal{M}, x \Vdash \Diamond\Diamond\top\} = W$. The next step is to prove that for every formula B which only contains the propositional variable p,

$$\mathcal{N}, x \Vdash' B \text{ iff } \mathcal{M}, x \Vdash B^\dagger \quad \text{(for every } x \in W\text{)}.$$

The proof of this claim proceeds by an induction on the length of B.

- For $\top$ or $\bot$ the claim is vacuous.
- If $B = p$ we have that $p^\dagger = \Diamond\Diamond\top \to \top \rhd \Diamond\top$. By construction, $\mathcal{M}, x \Vdash \Diamond\Diamond\top$. By construction, in any R-successor of x one can go by an S_x transition to some $e_i^\flat$ where $\Diamond\top$ holds. Thus, indeed, $\mathcal{M}, x \Vdash \top \rhd \Diamond\top$.
- On the other hand, if $\mathcal{N}, x \Vdash \neg p$, then again $\mathcal{M}, x \Vdash \Diamond\Diamond\top$. But in this case we can go via an R-transition to some $e_i^\natural$. By construction there is no S_x-transition from $e_i^\natural$ to any point where $\Diamond\top$ holds, whence $\mathcal{M}, x \nVdash \top \rhd \Diamond\top$.
- The proof of the claim is trivial for both the Boolean connectives and the modal operator $\Box$.

---

[5] W.r.t. the R-relation of course.

Now that the claim is established the lemma follows immediately.   □

4   Computational complexity of IL_0

First we obtain PSPACE hardness of **IL**.

4.1   PSPACE hardness

If we combine Lemma 3.1 and Lemma 3.2 we see that we have a reduction of $\mathbf{GL}_1$ to $\mathbf{IL}_0$. That is, for any formula A with at most one variable p we have that

$$\mathbf{GL} \vdash A \quad \Leftrightarrow \quad \mathbf{IL} \vdash A^\dagger.$$

As it is known that $\mathbf{GL}_1$ is PSPACE complete (see [3, Theorem 7] and [9]) we obtain the main result of this paper.

Theorem 4.1 *The computational complexity of $\mathbf{IL}_0$ is PSPACE hard.*

If in addition to this we would know that **IL** is in PSPACE we would obtain PSPACE completeness. It is commonly held that indeed the complexity of full **IL** is PSPACE-compleet, but up to now nobody has yet proven this. It came as a bit of a surprise to the authors to find out that actually no complexity results in the field of interpretability logics are known. Thus, this short note could well be the precursor to further investigations in various interpretability logics on very natural complexity questions in the otherwise mature field of interpretability logic.

4.2   On a normal form theorem for $\mathbf{IL}_0$

The PSPACE completeness of $\mathbf{IL}_0$ does a-priori not exclude the possibility of a normal form theorem of **IL**. It is even conceivable that there exists some easily recognizable class of normal forms for $\mathbf{IL}_0$ so that each formula in the language of $\mathbf{IL}_0$ is equivalent to a small sized boolean combination of these normal forms. In such a case the normal forms themselves may be easy and even easily comparable but then for an arbitrary formula it still remains hard (PSPACE) to see actually *what* combination of normal forms it is provably equivalent to. These observations render a normal form theorem for $\mathbf{IL}_0$ –if it would exist– useless for most practical purposes.

References

[1] G. Boolos. *The Logic of Provability*. Cambridge University Press, 1993.

[2] A. Chagrov and M. Zakharyaschev. *Modal Logic*, volume 35 of *Oxford Logic Guides*. Oxford University Press, 1997.

[3] A. V. Chagrov and M. N. Rybakov. How many variables does one needs to prove PSPACE-hardness of modal logics. volume 4 of *Advances in Modal Logic*, pages 71–82. 2003.

[4] D. de Jongh and G. Japaridze. The logic of provability. In S.R. Buss, editor, *Handbook of Proof Theory*, Studies in Logic and the Foundations of Mathematics, pages 475–546. Elsevier, Amsterdam, 1998.

[5] P. Hájek and V. Švejdar. A note on the normal form of closed formulas of interpretability logic. *Studia Logica*, 50(1):25–28, 1991.

[6] D. Leivant. On the proof theory of the modal logic for arithmetic provability. *Journal of Symbolic Logic*, 46(3):531–538, 1981.

[7] V. Čačić and M. Vukovič. A note on normal forms for closed fragment of system il. *Mathematical Communications*, To appear.

[8] A. Visser. An overview of interpretability logic. Advances in Modal Logic, pages 307–359. CSLI Publications, Stanford, CA, 1997.

[9] V. Švejdar. The decision problem of provability logic with only one atom. *Archive for Math. Logic*, 42(8):763–768, 2003.