The interpretability logic ILF∗

Mladen Vuković†

Abstract. In this paper we determine a characteristic class of $IL_{set}$-frames for the principle $F$. Then we prove that the principle $P$ is not provable in the system ILF. We use a generalized Veltman model.

Key words: interpretability logic, generalized Veltman semantic


Ključne riječi: logika interpretabilnosti, generalizirana Veltmanova semantika

AMS subject classifications: O3F25

Received October 5, 1998 Accepted October 26, 1998

1. Introduction

The interpretability logic $IL$ is the natural extension of provability logic. The language of the interpretability logic contains propositional letters $p_0, p_1, \ldots$, the logical connectives $\land, \lor, \to, \neg$, and the unary modal operator $\square$ and the binary modal operator $\triangleright$. We use $\bot$ for false and $\top$ for true. The axioms of the interpretability logic $IL$ are:

(L0) all tautologies of the propositional calculus

(L1) $\square(A \to B) \to (\square A \to \square B)$

(L2) $\square A \to \square \square A$

(L3) $\square(\square A \to A) \to \square A$

(J1) $\square(A \to B) \to (A \triangleright B)$

(J2) $(A \triangleright B \land B \triangleright C) \to (A \triangleright C)$

(J3) $((A \triangleright C) \land (B \triangleright C)) \to ((A \lor B) \triangleright C)$

∗This paper is part of the lecture presented at the Mathematical Colloquium in Osijek organized by Croatian Mathematical Society - Division Osijek, May 29, 1998.

†Department of Mathematics, University of Zagreb, Bijenička 30, HR-10 000 Zagreb, e-mail: vukovic@zrcmath.math.hr
(J4) \((A \Rightarrow B) \rightarrow (\Diamond A \rightarrow \Diamond B)\)

(J5) \(\Diamond A \Rightarrow A\)

where \(\Diamond\) stands for \(\neg \Box \neg\) and \(\Rightarrow\) has the same priority as \(\rightarrow\). The deduction rules of \(IL\) are modus ponens and necessitation.

Various extensions of \(IL\) are obtained by adding some new axioms. These new axioms are called the principles of interpretability. We observe here the principle \(P : A \Rightarrow B \rightarrow \Box (A \Rightarrow B)\) (principle of persistence) and \(F : (A \Rightarrow \Diamond A) \rightarrow \Box (\neg A)\) (Feferman's principle).

In this paper we determine a characteristic class of \(IL_{\text{set}}\)-frames for the principle \(F\). Then we prove independence of the principle \(P\) in the system \(ILF\).

2. The Generalized Veltman semantic

Now we define the generalized Veltman semantic for the interpretability logic.

**Definition 1. (de Jongh)** An ordered triple \((W, R, \{S_w : w \in W\})\) is called the \(IL_{\text{set}}\)-frame, and denoted by \(W\), if we have:

a) \((W, R)\) is a \(L\)-frame, i.e. \(W\) is a non-empty set, and \(R\) is a transitive and reverse well-founded relation on \(W\) (the elements of \(W\) we call nodes);

b) Every \(w \in W\) satisfies

\[
S_w \subseteq W[w] \times P(W[w]) \setminus \{\emptyset\},
\]

where \(W[w]\) denotes the set \(\{x : wRx\}\);

c) The relation \(S_w\) is quasi-reflexive for every \(w \in W\), i.e. \(wRx\) implies \(xS_w\{x\}\);

d) The relation \(S_w\) is quasi-transitive for every \(w \in W\), i.e. if \(xS_wY\) and \((\forall y \in Y)(yS_wZ_y)\) then \(xS_w(\bigcup y \in Y Z_y)\);

e) If \(wRuRv\) then \(uS_w\{v\}\);

f) If \(xS_wY\) and \(Y \subseteq Z \subseteq W[w]\) then \(xS_wZ\).

**Definition 2. (de Jongh)** An ordered quadruple \((W, R, \{S_w : w \in W\}, \vdash)\) is called the \(IL_{\text{set}}\)-model (generalized Veltman model), and denoted by \(W\), if we have:

(1) \((W, R, \{S_w : w \in W\})\) is an \(IL_{\text{set}}\)-frame;

(2) \(\vdash\) is the forcing relation between elements of \(W\) and formulas of \(IL\), which satisfies the following:

(2a) \(w \vdash \top\) and \(w \not\vdash \bot\) are valid for every \(w \in W\);

(2b) \(\vdash\) commutes with the Boolean connectives;

(2c) \(w \vdash \Box A\) if and only if \(\forall x(wRx \Rightarrow x \vdash A)\);

(2d) \(w \vdash A \Rightarrow B\) if and only if

\[
\forall v((wRv \land v \vdash A) \Rightarrow \exists V(vS_wV \land (\forall x \in V)(x \vdash B))).
\]
As usual we shall use the same letter $W$ for a model and a frame. If $W$ is an IL$_{set}$-frame and $A$ is a formula of $IL$, we write $W \models A$ iff $w \vdash A$ for all forcing relations $\vdash$ on $W$ and all nodes $w$ of $W$.

For a modal scheme $(A)$ and an IL$_{set}$-frame $W$, $W \models (A)$ denotes the fact that $W \models B$ for an arbitrary instance $B$ of $(A)$. Analogously, we define $W \models A$ and $W \models (A)$, if $W$ is an IL$_{set}$-model. If $W$ is an IL$_{set}$-model, $v \vdash A$ means that $v \vdash A$ for any $v \in V$.

It is easy to check the adequacy of the system $IL$ with respect to IL$_{set}$-models. In [6] we proved the completeness of the system $IL$ with respect to generalized Veltman models.

Let $\Gamma$ be a set of modal formulas. We will say that an IL$_{set}$-frame $W= (W, R, \{S_w : w \in W\})$ is in the characteristic class of $\Gamma$ if we have $W \models \Gamma$, for all forcing relations $\vdash$ on $W$. The characteristic class of a principle of interpretability is the characteristic class of the set of all instances of the principle. By $(A)^*$ we denote a property of an IL$_{set}$-frame which determines the characteristic class of some principle $A$.

R. Verbrugge determined in [2] the characteristic classes of the principle $P$. Denote by $(P)^*$ the following property of an IL$_{set}$-frame:

$$x_3S_{x_2}Y \& x_1R_{x_2}Rx_3 \Rightarrow (\exists Y' \subseteq Y)(x_3S_{x_2}Y').$$

3. The system $ILF$

S. Feferman proved the generalization of Gödel’s second incompleteness theorem, i.e. the formula $Cons$ which expresses the consistency of Peano arithmetic) is not interpretable in $P\ A$. The Feferman’s principle $F : (A \triangleright \Box A) \rightarrow \Box(\neg A)$ is a modal description of Feferman’s theorem.

V. Švejdar in [1] proved $IL(KW1^\circ) \vdash F$ and $ILW \vdash KW1^\circ$. We proved in [7] (Corollary 5.16) that $ILW \nvdash P$. Švejdar’s and our results imply $ILF \nvdash P$. In Proposition 3 we will prove the same result more directly (without using Švejdar’s result).

V. Švejdar determined a characteristic class of (ordinary) Veltman’s frames for the principle $F$. His proofs of independences in system $ILF$ are relatively complicated. A problem is that principles $F, W, KW1^\circ$ have the same characteristic classes. In [7] we proved that the principle $F, W, KW1^\circ$ have different characteristic class of IL$_{set}$-frames. So we have simpler proofs of independences than Švejdar.

By the following definition we give relations which we use for the characteristic class of IL$_{set}$-frames for the principle $F$.

**Definition 3.** Let $(W, R, \{S_w : w \in W\})$ be IL$_{set}$-frame and $w \in W$. We denote with $S_w$ and $R_{w}$ the following relations:

$$KW1 : (A \triangleright \Box T) \rightarrow (T \triangleright (\neg A)), \ KW1^\circ : ((A \& B) \triangleright \Box A) \rightarrow (A \triangleright (A \& (\neg B))), \ W : (A \triangleright B) \rightarrow (A \triangleright (B \& \Box(\neg A))).$$
for $\emptyset \neq A \subseteq W[w]$ and $B \subseteq \mathcal{P}(W[w]) \backslash \{\emptyset\}$ is valid

$$A \bar{S}_w B \iff (\forall a \in A)(\exists B \in B)(aS_w B);$$

for $C \subseteq \mathcal{P}(W[w]) \backslash \{\emptyset\}$ and $\emptyset \neq D \subseteq W[w]$ is valid

$$C \bar{R}_w D \iff (\forall C \in C)(\forall c \in C)(\exists d \in D)(cR_d).$$

We denote by $(F)^*$ the following property of an $\text{IL}_{\text{set}}$-frame:

$$\text{relation } \bar{S}_w \circ \bar{R}_w \text{ is reverse well-founded for all } w \in W.$$ 

**Proposition 1.** Let $W$ be an $\text{IL}_{\text{set}}$-frame. We have

$$W \models F \iff \text{ if and only if } W \text{ satisfies }(F)^*$$

**Proof.** Let us suppose that the frame $W$ does not have the property $(F)^*$, i.e. there is a node $w \in W$ such that relation $\bar{S}_w \circ \bar{R}_w$ is not reverse well-founded. So there are sequences of sets $A_1 , A_2 , \ldots$ and $B_1 , B_2 , \ldots$ such that

$$A_1 \bar{S}_w B_1 \bar{R}_w A_2 \bar{S}_w B_2 \ldots$$

Now we define a forcing relation $\vdash$ on $W$ by:

$$a \vdash p \iff a \in \bigcup_{i=1}^{\infty} A_i .$$

We claim that $w \not\vdash (p \triangledown \phi) \implies \square (\neg p)$, because $wR_a$ and $a \vdash \phi$ for all $a \in A_1$. We have $w \not\vdash \square (\neg p)$ because $wR_a$ and $a \vdash \phi$ for all $a \in A_1$. The claim $w \not\vdash \phi \implies \square \phi$ is equivalent to

$$\forall x(wRx \& x \vdash p \implies \exists Y (xS_w Y \& (\forall y \in Y)(\exists z)(yRz \& z \vdash \phi))).$$

Let $x \in W$ is such that $wRx$ and $x \vdash \phi$. By definition of the relation $\vdash$ there is $i \in \mathbb{N}$ such that $x \in A_i$. By definition of the relation $\bar{S}_w$, and facts $A_i , \bar{S}_w B_i$ and $x \in A_i$ there is $Y \in B_i$ such that $x \bar{S}_w Y$. By $B_i \bar{R}_w A_{i+1}$ and $Y \in B_i$ we have $(\forall y \in Y)(\exists z \in A_{i+1})(yRz)$. The fact $z \in A_{i+1}$ implies $z \vdash \phi$. So we proved $w \not\vdash \phi \implies \square \phi$.

Now, we prove the condition $(F)^*$ is sufficient for the principle $F$. Let $\text{IL}_{\text{set}}$-frame $W$ satisfy the condition $(F)^*$, and let $\vdash$ be a forcing relation on $W$. Let $w \in W$ be such that $w \not\vdash A \triangledown \phi$, i.e.

$$\forall x(wRx \& x \vdash A) \implies \exists Y (xS_w Y \& (\forall y \in Y)(\exists z)(yRz \& z \vdash A))) \quad (*)$$

Now we suppose that there is $x_1 \in W$ such that $wRx_1$ and $x_1 \vdash A$. By $(*)$ there is $Y_1 \subseteq W[w]$ such that $x_1S_w Y_1$ and

$$(\forall y \in Y_1)(\exists z_y)(yRz_y \& z_y \not\vdash A).$$

So the facts $\{x_1\}S_w \{Y_1\}$ and $\{Y_1\}R_w \{z_y\} : y \in Y_1$ are true. From this we have

$$(\forall y \in Y_1)\{x_1\}S_w \{Y_1\} \bar{R}_w \{z_y\} : y \in Y_1.$$
For all nodes \( z^{(1)}_y \) we have \( wRz^{(1)}_y \) and \( z^{(1)}_y \vdash A \). Then the fact \((*)\) implies that for all \( y \in Y_1 \) there is \( Z_{2,y} \subseteq W[w] \) such that \( z^{(1)}_y S_w Y_{2,y} \) and

\[
(\forall u \in Y_{2,y})(\exists z^{(2)}_{y,u})(uRz^{(2)}_{y,u}) \& z^{(2)}_{y,u} \vdash A.
\]

So we have

\[
\{ Y_{2,y} : y \in Y_1 \} \overline{R_w} \{ z^{(2)}_{y,u} : y \in Y_1, u \in Y_{2,y} \}.
\]

Also we proved

\[
\{ x_1 \} (\overline{S_w} \circ \overline{R_w}) \{ z^{(1)}_y : y \in Y_1 \} (\overline{S_w} \circ \overline{R_w}) \{ z^{(2)}_{y,u} : y \in Y_1, u \in Y_{2,y} \},
\]

and

\[
(\forall y \in Y_1)(\forall u \in Y_{2,y})(z^{(2)}_{y,u} \vdash A).
\]

From this we conclude that the fact \((*)\) can be used again. Also, the last construction can be repeated infinitely many times. So the relation \( \overline{S_w} \circ \overline{R_w} \) is not reverse well-founded, what is a contradiction. This means that \( w \vdash \Box(\neg A) \), i.e. \( w \vdash \Box F \). \( \square \)

**Proposition 2.** We have \( \text{ILF} \not\vdash P \).

**Proof.** By the following picture we give \( \text{IL}_{\text{set}} \)-frame \( W \).

![Diagram](image)

Full arrows in the picture indicate the relation \( R \), while the dotted ones indicate \( S_w \). The relations between nodes (transitivity of the relation \( R; wRvRu \Rightarrow vS_w\{u\}; \) quasi-reflexivity and quasi-transitivity of \( S_w \); condition f) in the definition of \( \text{IL}_{\text{set}} \)-frame) will not be indicated by arrows.

In the picture we have \( wRvRb \) and \( bS_w\{a\} \) but \( bS_w\{a\} \) is not valid. So the \( \text{IL}_{\text{set}} \)-frame does not have the property \((P)^*\).

It is easy to see that \( \overline{S_x} \circ \overline{R_x} \) is reverse well-founded relation for all \( x \in W \). \( \square \)

**References**


