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*

Sažetak. **Sistem ILF za logiku interpretabilnosti.** *U ovom članku odredili smo karakterističnu klasu IL_{skup}-okvira za princip F. Pomoću toga dokazujemo da princip P nije dokaziv u sistemu ILF. U dokazu koristimo generalizirane Veltmanove modele.*

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 $\wedge,\ \vee, \rightarrow,\ \neg$, and the unary modal operator $\Box$ and the binary modal operator $\rhd$. 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) $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

(L2) $\Box A \rightarrow \Box\Box A$

(L3) $\Box(\Box A \rightarrow A) \rightarrow \Box A$

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

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

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

*Thih 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@cromath.math.hr

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

(J5) $\Diamond A \rhd A$

where $\Diamond$ stands for $\neg\Box\neg$ and $\rhd$ has the same priority as $\to$. 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 $P\ :\ A \rhd B \to \Box(\mathrm{A} \rhd B)$ (principle of persistence) and $F\ :\ (A \rhd \Diamond A) \to \Box(\neg A)$ (Feferman's principle).

In this paper we determine a characteristic class of IL_{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_{set}*-frame, and denoted by* $\boldsymbol{W}$*, 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*

$$S_w \subseteq W[w] \times \mathcal{P}(W[w])\backslash\{\emptyset\}\ ,$$

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* $xS_w(\cup_{y\in Y}Z_y)$;

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_{set}*-model (generalized Veltman model), and denoted by* $\boldsymbol{W}$*, if we have:*

(1) $(W, R, \{S_w : w \in W\})$ *is an* IL_{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* $w \nVdash \bot$ *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) $w \Vdash A \rhd B$ *if and only if*

$$\forall v((wRv\ \&\ v \Vdash A)\ \Rightarrow\ \exists V(vS_wV\ \&\ (\forall x \in V)(x \Vdash B))).$$

As usual we shall use the same letter $\boldsymbol{W}$ for a model and a frame. If $\boldsymbol{W}$ is an IL_{set}-frame and A is a formula of IL, we write $\boldsymbol{W} \models A$ iff $w \Vdash A$ for all forcing relations $\Vdash$ on $\boldsymbol{W}$ and all nodes w of W.

For a modal scheme (A) and an IL_{set}-frame $\boldsymbol{W}$, $\boldsymbol{W} \models (A)$ denotes the fact that $\boldsymbol{W} \models B$ for an arbitrary instance B of (A). Analogously, we define $\boldsymbol{W} \models A$ and $\boldsymbol{W} \models (A)$, if $\boldsymbol{W}$ is an IL_{set}-model. If $\boldsymbol{W}$ is an IL_{set}-model, $V \subseteq W$ and A a formula, the notation $V \Vdash A$ 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 Γ be a set of modal formulas. We will say that an IL_{set}-frame $\boldsymbol{W}=(W, R, \{S_w \ : \ w \in W\})$ is in the characteristic class of Γ if we have $W \models \Gamma$, for all forcing relations $\Vdash$ on $\boldsymbol{W}$. 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 :

$$x_3 S_{x_1} Y \ \& \ x_1 R x_2 R x_3 \ \Rightarrow \ (\exists Y' \subseteq Y)(x_3 S_{x_2} Y').$$

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 PA. The Feferman's principle $F \ : \ (A \rhd \Diamond A) \to \Box(\neg A)$ is a modal description od 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 $ILW \nvdash P$. [1] 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 $\overline{S_w}$ and $\overline{R_w}$ the following relations:*

[1] $KW1 \ : \ (A \rhd \Diamond\top) \to (\top \rhd (\neg A))$, $KW1^\circ \ : \ ((A \wedge B) \rhd \Diamond A) \to (A \rhd (A \wedge (\neg B)))$, $W \ : \ (A \rhd B) \to (A \rhd (B \wedge \Box(\neg A)))$.

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

$$A\overline{S_w}\mathcal{B} \;\Leftrightarrow\; (\forall a \in A)(\exists B \in \mathcal{B})(aS_wB);$$

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

$$\mathcal{C}\overline{R_w}D \;\Leftrightarrow\; (\forall C \in \mathcal{C})(\forall c \in C)(\exists d \in D)(cRd)\ .$$

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

relation $\overline{S_w} \circ \overline{R_w}$ *is reverse well-founded for all* $w \in W$.

Proposition 1. *Let* $\boldsymbol{W}$ *be an* IL_{set}*-frame. We have*

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

Proof. Let us suppose that the frame $\boldsymbol{W}$ does not have the property $(F)^*$, i.e. there is a node $w \in W$ such that relation $\overline{S_w} \circ \overline{R_w}$ is not reverse well-founded. So there are sequences of sets A_1 , A_2 , … and $\mathcal{B}_1$, $\mathcal{B}_2$, … such that

$$A_1\ \overline{S_w}\ \mathcal{B}_1\ \overline{R_w}\ A_2\ \overline{S_w}\ \mathcal{B}_2\ \ldots$$

Now we define a forcing relation $\Vdash$ on $\boldsymbol{W}$ by:

$$a \Vdash p \;\Leftrightarrow\; a \in \bigcup_{i=1}^{\infty} A_i\ .$$

We claim that $w \nVdash (p \rhd \Diamond p) \rightarrow \Box(\neg p)$. We have $w \nVdash \Box(\neg p)$, because wRa and $a \Vdash p$ for all $a \in A_1$. The claim $w \Vdash p \rhd \Diamond p$ is equivalent to

$$\forall x(wRx\ \&\ x \Vdash p\ \Rightarrow \exists Y(xS_wY\ \&\ (\forall y \in Y)(\exists z)(yRz\ \&\ z \Vdash p))).$$

Let $x \in W$ is such that wRx and $x \Vdash p$. By definition of the relation $\Vdash$ there is $i \in \boldsymbol{N}$ such that $x \in A_i$. By definition of the relation $\overline{S_w}$, and facts $A_i\overline{S_w}\mathcal{B}_i$ and $x \in A_i$ there is $Y \in \mathcal{B}_i$ such that $x\overline{S_w}Y$. By $\mathcal{B}_i\overline{R_w}A_{i+1}$ and $Y \in \mathcal{B}_i$ we have $(\forall y \in Y)(\exists z \in A_{i+1})(yRz)$. The fact $z \in A_{i+1}$ implies $z \Vdash p$. So we proved $w \Vdash p \rhd \Diamond p$.

Now, we prove that the condition $(F)^*$ is sufficient for the principle F. Let IL_{set}-frame $\boldsymbol{W}$ satisfy the condition $(F)^*$, and let $\Vdash$ be a forcing relation on $\boldsymbol{W}$. Let $w \in W$ be such that $w \Vdash A \rhd \Diamond A$, i.e.

$$\forall x((wRx\ \&\ x \Vdash A)\ \Rightarrow\ \exists Y(xS_wY\ \&\ (\forall y \in Y)(\exists z)(yRz\ \&\ z \Vdash A))) \qquad (*)$$

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 $x_1S_wY_1$ and

$$(\forall y \in Y_1)(\exists z_y^{(1)})(yRz_y^{(1)}\ \&\ z_y^{(1)} \Vdash A).$$

So the facts $\{x_1\}\overline{S_w}\{Y_1\}$ and $\{Y_1\}\overline{R_w}\{z_y^{(1)}\ :\ y \in Y_1\}$ are true. From this we have

$$\{x_1\}(\overline{S_w} \circ \overline{R_w})\{z_y^{(1)}\ :\ y \in Y_1\}.$$

For all nodes $z_y^{(1)}$ we have $wRz_y^{(1)}$ and $z_y^{(1)} \Vdash A$. Then the fact $(*)$ implies that for all $y \in Y_1$ there is $Y_{2,y} \subseteq W[w]$ such that $z_y^{(1)} S_w Y_{2,y}$ and

$$(\forall u \in Y_{2,y})(\exists z_{y,u}^{(2)})(uRz_{y,u}^{(2)}) \ \& \ z_{y,u}^{(2)} \Vdash A).$$

So we have

$$\{Y_{2,y} \ : \ y \in Y_1\}\overline{R_w}\{z_{y,u}^{(2)} \ : \ y \in Y_1, \ u \in Y_{2,y}\}.$$

Also we proved

$$\{x_1\}(\overline{S_w} \circ \overline{R_w})\{z_y^{(1)} \ : \ y \in Y_1\}(\overline{S_w} \circ \overline{R_w})\{z_{y,u}^{(2)} \ : \ y \in Y_1, \ u \in Y_{2,y}\},$$

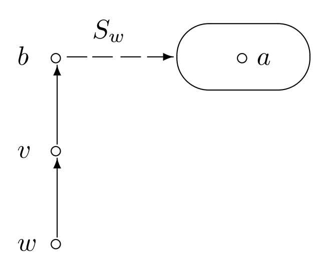
and

$$(\forall y \in Y_1)(\forall u \in Y_{2,y})(z_{y,u}^{(2)} \Vdash A).$$

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-fonded, what is a contradiction. This means that $w \Vdash \Box(-A)$, i.e. $w \Vdash F$. $\Box$

Proposition 2. *We have $ILF \nvdash P$.*

Proof. By the following picture we give IL_{set}-frame W.



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 IL_{set}-frame) will not be indicated by arrows.

In the picture we have $wRvRb$ and $bS_w\{a\}$ but $bS_v\{a\}$ is not valid. So the IL_{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 $x \in W$. $\Box$

References

[1] V. ŠVEJDAR, *Some Independence Results in Interpretability Logic*, Studia Logica **50**(1991), 29–38.

[2] R. VERBRUGGE, *Generalized Veltman frames and models*, manuscript

[3] A. Visser, *Interpretability logic*, In: Math. Logic, Proceedings of the 1988 Heyting Conference, (P. P. Petkov, Ed.), Plenum Press, New York, 1990, 175–210.

[4] A. Visser, *The Formalization of Interpretability*, Studia Logica **50**(1991), 81–105.

[5] A. Visser, *An overview of Interpretability Logic*, Logic group preprint series, no. 174, Department of Philosophy, University of Utrecht, 1997.

[6] M. Vuković, *Some Correspondences of Principles in Interpretability Logic*, Glasnik Matematički **31**(**51**) (1996), 193–200.

[7] M. Vuković, *Generalized Veltman models*, Ph. D. Thesis, University of Zagreb, 1996 (in Croatian).

[8] M. Vuković, *The principles of interpretability*, submitted