Interpretability in PRA

Marta Bilkova†, Dick de Jongh*, and Joost J. Joosten*,

*Institute for Logic Language and Computation
University of Amsterdam
and
†Department of Logic
Charles University; Prague

14th July 2007
We all use the notion $T 	riangleright S$: $T$ interprets $S$
We all use the notion $T hd S$: $T$ interprets $S$

$T hd S$ means (modulo some technical details)
We all use the notion $T \triangleright S$: $T$ interprets $S$

$T \triangleright S$ means (modulo some technical details)

$\exists j \forall \varphi (\text{Axiom}_S(\varphi) \rightarrow \exists p \ \text{Proof}_T(p, \lceil \varphi^j \rceil))$
We are interested in the structural behavior of the notion of interpretability.
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.

We shall consider sentential extensions of a base theory.
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.

We shall consider sentential extensions of a base theory

$\varphi \triangleright T \psi$ stands for
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.

We shall consider sentential extensions of a base theory.

\( \varphi \triangleright T \psi \) stands for

\( T + \varphi \triangleright T + \psi \)
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.

We shall consider sentential extensions of a base theory

\( \varphi \triangleright_T \psi \) stands for

\( T + \varphi \triangleright T + \psi \)

We are interested in the interpretability logic of a theory \( T \):
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmeticized.

We shall consider sentential extensions of a base theory

\( \varphi \vdash T \psi \) stands for

\( T + \varphi \vdash T + \psi \)

We are interested in the interpretability logic of a theory \( T \):

The set of all model propositional logical formulas in the language \( \Box, \vdash \) which are true regardless how you interpret the variables as arithmetical sentences.
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.

We shall consider sentential extensions of a base theory

\( \varphi \triangleright T \psi \) stands for

\( T + \varphi \triangleright T + \psi \)

We are interested in the interpretability logic of a theory \( T \):

The set of all model propositional logical formulas in the language \( \Box, \triangleright \) which are true regardless how you interpret the variables as arithmetical sentences

Of course, reading \( \triangleright \) as \( \triangleright_T \), etc.
We are interested in the structural behavior of the notion of interpretability.

Interpretability can easily be formalized/arithmetized.

We shall consider sentential extensions of a base theory

\[ \varphi \triangleright_T \psi \text{ stands for} \]

\[ T + \varphi \triangleright T + \psi \]

We are interested in the interpretability logic of a theory \( T \):

The set of all model propositional logical formulas in the language \( \Box, \triangleright \) which are true regardless how you interpret the variables as arithmetical sentences

Of course, reading \( \triangleright \) as \( \triangleright_T \), etc.

Example: \( (\varphi \triangleright \psi) \land (\psi \triangleright \chi) \rightarrow (\varphi \triangleright \chi) \)
For all theories $T$, $\text{IL}(T)$ contains some sort of core logic $\text{IL}$
For all theories $T$, $\text{IL}(T)$ contains some sort of core logic $\text{IL}$

$\text{IL}(T)$ is characterized by some additional axiom schemes on top of that
For all theories \( T \), \( \text{IL}(T) \) contains some sort of core logic IL

\( \text{IL}(T) \) is characterized by some additional axiom schemes on top of that

For example, for theories with full induction, we have that *Montagna's Axiom* holds

\[
(A \Rightarrow B) \rightarrow ((A \land \Box C) \Rightarrow (B \land \Box C))
\]
For all theories \( T \), \( IL(T) \) contains some sort of core logic IL

\( IL(T) \) is characterized by some additional axiom schemes on top of that

For example, for theories with full induction, we have that _Montagna’s Axiom_ holds

\[
(A \triangleright B) \rightarrow ((A \land 
\square C) \triangleright (B \land 
\square C))
\]

It turns out that precisely ILM is, e.g. IL(PA) (Shavrukov 1988; Berarducci 1990)
For all theories $T$, $\text{IL}(T)$ contains some sort of core logic $\text{IL}$

$\text{IL}(T)$ is characterized by some additional axiom schemes on top of that

For example, for theories with full induction, we have that Montagna's Axiom holds

$$(A \triangleright B) \rightarrow ((A \land \Box C) \triangleright (B \land \Box C))$$

It turns out that precisely ILM is, e.g. $\text{IL}(\text{PA})$ (Shavrukov 1988; Berarducci 1990)

Likewise, the interpretability logic for finitely axiomatized theories is known
For all theories $T$, $IL(T)$ contains some sort of core logic $IL$

$IL(T)$ is characterized by some additional axiom schemes on top of that

For example, for theories with full induction, we have that

$Montagna's Axiom$ holds

$$(A \triangleright B) \rightarrow ((A \land \Box C) \triangleright (B \land \Box C))$$

It turns out that precisely $ILM$ is, e.g. $IL(PA)$ (Shavrukov 1988; Berarducci 1990)

Likewise, the interpretability logic for finitely axiomatized theories is known

And no other!
For all theories $T$, $IL(T)$ contains some sort of core logic $IL$

$IL(T)$ is characterized by some additional axiom schemes on top of that

For example, for theories with full induction, we have that Montagna's Axiom holds

$$(A 	riangleright B) \rightarrow ((A \land \Box C) \triangleright (B \land \Box C))$$

It turns out that precisely ILM is, e.g. $IL(PA)$ (Shavrukov 1988; Berarducci 1990)

Likewise, the interpretability logic for finitely axiomatized theories is known

And no other!

That's were PRA comes in
Consider again

\[ \exists j \, \forall \varphi (\text{Axiom}_S(\varphi) \rightarrow \exists p \, \text{Proof}_T(p, \varphi^j)) \]
Consider again the formula:

$$\exists j \ \forall \varphi (\text{Axiom}_S(\varphi) \rightarrow \exists p \ \text{Proof}_T(p, \neg \varphi^j \neg))$$

Certainly $\Sigma_3$
Consider again

\[ \exists j \forall \varphi (\text{Axiom}_S(\varphi) \rightarrow \exists p \ \text{Proof}_T(p, \neg \varphi^j)) \]

Certainly \( \Sigma_3 \)

When \( S \) has finitely many axioms, then \( \Sigma_1 \)
Consider again

$$\exists j \ \forall \varphi (\text{Axiom}_S(\varphi) \rightarrow \exists p \ \text{Proof}_T(p, \square \varphi^j \land))$$

Certainly $\Sigma_3$

When $S$ has finitely many axioms, then $\Sigma_1$

When $T$ is reflexive, then $\Pi_2$. (Orey-Hájek).
Consider again

\[ \exists j \ \forall \varphi ( \text{Axiom}_S(\varphi) \rightarrow \exists p \ \text{Proof}_T(p, \neg \varphi^j \neg)) \]

Certainly \( \Sigma_3 \)

When \( S \) has finitely many axioms, then \( \Sigma_1 \)

When \( T \) is reflexive, then \( \Pi_2 \). (Orey-Hájek).

When \( T \) is reflexive, we have access to Montagna’s Principle:

\[ (T \triangleright S) \rightarrow ((T \land \Box \gamma) \triangleright (S \land \Box \gamma)) \]
Consider again

\[ \exists j \, \forall \varphi (\text{Axiom}_S (\varphi) \rightarrow \exists p \, \text{Proof}_T (p, \neg \varphi^j^\bot)) \]

Certainly \( \Sigma_3 \)

When \( S \) has finitely many axioms, then \( \Sigma_1 \)

When \( T \) is reflexive, then \( \Pi_2 \). (Orey-Hájek).

When \( T \) is reflexive, we have access to Montagna’s Principle:

\[(T \triangleright S) \rightarrow ((T \land \Box \gamma) \triangleright (S \land \Box \gamma)) \]

Every extension of PRA by \( \Sigma_2 \) sentences is reflexive (Parsons, Beklemishev, etc)
Consider again

$$\exists j \ \forall \varphi (\text{Axiom}_S(\varphi) \rightarrow \exists p \ \text{Proof}_T(p, \neg \varphi^j \upharpoonright))$$

Certainly $\Sigma_3$

When $S$ has finitely many axioms, then $\Sigma_1$

When $T$ is reflexive, then $\Pi_2$. (Orey-Hájek).

When $T$ is reflexive, we have access to Montagnà’s Principle:

$$(T \triangleright S) \rightarrow ((T \land \Box \gamma) \triangleright (S \land \Box \gamma))$$

Every extension of PRA by $\Sigma_2$ sentences is reflexive (Parsons, Beklemishev, etc)

$$(\alpha \triangleright_{\text{PRA}} \beta) \rightarrow ((\alpha \land \Box \gamma) \triangleright_{\text{PRA}} (\beta \land \Box \gamma))$$

whenever $\alpha \in \Sigma_2$
$B := (A \triangleright B) \rightarrow (A \land \Box C) \triangleright (B \land \Box C)$ for $A \in \text{ES}_2$
\[ B := (A \triangleright B) \rightarrow (A \land \Box C) \triangleright (B \land \Box C) \quad \text{for } A \in ES_2 \]

where
\[ B := (A \triangleright B) \rightarrow (A \land \Box C) \triangleright (B \land \Box C) \quad \text{for } A \in ES_2 \]

where

\[ ES_2 := \Box A \mid \neg \Box A \mid ES_2 \land ES_2 \mid ES_2 \lor ES_2 \mid \neg (ES_2 \triangleright A) \]
If $T$ and $S$ are $\Pi_2$ axiomatized theories with
If $T$ and $S$ are $\Pi_2$ axiomatized theories with $T \equiv_1 S$. 

In other words: $\alpha, \beta \in \Delta_2^\Pi$, and $\alpha, \beta \in \Pi_2$. 

Interpretability in PRA
If $T$ and $S$ are $\Pi_2$ axiomatized theories with $T \equiv_1 S$, then, $T \equiv_1 (T \cup S)$.
If $T$ and $S$ are $\Pi_2$ axiomatized theories with

1. $T \equiv_1 S$
2. then, $T \equiv_1 (T \cup S)$

So,

$$(\alpha \triangleright \beta) \wedge (\beta \triangleright \alpha) \rightarrow (\alpha \triangleright (\alpha \wedge \beta))$$

whenever,
If $T$ and $S$ are $\Pi_2$ axiomatized theories with

- $T \equiv_1 S$
- then, $T \equiv_1 (T \cup S)$
- So,

\[(\alpha \triangleright \beta) \land (\beta \triangleright \alpha) \rightarrow (\alpha \triangleright (\alpha \land \beta))\]

whenever,

- $\alpha, \beta \in \Sigma_2$, and
If $T$ and $S$ are $\Pi_2$ axiomatized theories with

- $T \equiv_1 S$
- then, $T \equiv_1 (T \cup S)$
- So,

\[(\alpha \triangleright \beta) \land (\beta \triangleright \alpha) \rightarrow (\alpha \triangleright (\alpha \land \beta))\]

whenever,

- $\alpha, \beta \in \Sigma_2$, and
- $\alpha, \beta \in \Pi_2$. 
If $T$ and $S$ are $\Pi_2$ axiomatized theories with

$T \equiv_1 S$

then, $T \equiv_1 (T \cup S)$

So,

$$(\alpha \triangleright \beta) \land (\beta \triangleright \alpha) \rightarrow (\alpha \triangleright (\alpha \land \beta))$$

whenever,

$\alpha, \beta \in \Sigma_2$, and

$\alpha, \beta \in \Pi_2$.

In other words: $\alpha, \beta \in \Delta_2$
$Z \ (A \triangleright B) \land (B \triangleright A) \rightarrow (A \triangleright (A \land B))$ for $A$ and $B$ in $\text{ED}_2$
\[ Z \ (A \triangleright B) \land (B \triangleright A) \rightarrow (A \triangleright (A \land B)) \] for \( A \) and \( B \) in \( \text{ED}_2 \)

\[ \text{ED}_2 \ := \ \Box A \mid \neg \text{ED}_2 \mid \text{ED}_2 \land \text{ED}_2 \mid \text{ED}_2 \lor \text{ED}_2 \]
Why and how study interpretability

Proof theoretic characteristics of PRA

Modal matters

Beklemishev’s principle

Zambella’s Principle

\[ Z \ (A \triangleright B) \land (B \triangleright A) \rightarrow (A \triangleright (A \land B)) \] for \( A \) and \( B \) in ED\(_2\)

\[ \text{ED}_2 := \Box A \mid \neg \text{ED}_2 \mid \text{ED}_2 \land \text{ED}_2 \mid \text{ED}_2 \lor \text{ED}_2 \]

Is this all?
The logic $\mathcal{IL}$

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: $\Box (A \rightarrow B) \rightarrow A \triangleright B$
J2: $(A \triangleright B) \land (B \triangleright C) \rightarrow A \triangleright C$
J3: $(A \triangleright C) \land (B \triangleright C) \rightarrow A \lor B \triangleright C$
J4: $A \triangleright B \rightarrow (\Diamond A \rightarrow \Diamond B)$
J5: $\Diamond A \triangleright A$
A Veltman frame $F = \langle W, R, S \rangle$, where:
- $R \subseteq W \times W$,
- $S_w \subseteq W \times W$ for each $w \in W$. 

Frame conditions:
- $R$ is conversely well-founded and transitive,
- $ySx \rightarrow xRy \land xRz \rightarrow ySxz$, and
- $S_x$ is transitive and reflexive for each $x$. 

Marta Bilkova†, Dick de Jongh*, and Joost J. Joosten*, Interpretability in PRA
A Veltman frame $F = \langle W, R, S \rangle$,
- $R \subseteq W \times W$,
- $S_w \subseteq W \times W$ for each $w \in W$.

- $R$ is conversely well-founded and transitive.
A Veltman frame $F = \langle W, R, S \rangle$,
- $R \subseteq W \times W$,
- $S_w \subseteq W \times W$ for each $w \in W$.
- $R$ is conversely well-founded and transitive
- $yS_x z \rightarrow xRy \land xRz$
A Veltman frame $F = \langle W, R, S \rangle$, 
$R \subseteq W \times W$, 
$S_w \subseteq W \times W$ for each $w \in W$.

$R$ is conversely well-founded and transitive

$yS_x z \rightarrow xRy \land xRz$

$xRyRz \rightarrow yS_x z$
A Veltman frame $F = \langle W, R, S \rangle$, 
$R \subseteq W \times W$, 
$S_w \subseteq W \times W$ for each $w \in W$.

- $R$ is conversely well-founded and transitive
- $yS_xz \rightarrow xRy \land xRz$
- $xRyRz \rightarrow yS_xz$
- $S_x$ is transitive and reflexive for each $x$
A Veltman frame \( F = \langle W, R, S \rangle \),
\( R \subseteq W \times W \),
\( S_w \subseteq W \times W \) for each \( w \in W \).

- \( R \) is conversely well-founded and transitive
- \( yS_xz \rightarrow xRy \land xRz \)
- \( xRyRz \rightarrow yS_xz \)
- \( S_x \) is transitive and reflexive for each \( x \)
A model $M = \langle W, R, S, \models \rangle$,
$\models \subseteq W \times \text{Prop}$

$\upmodels w \not\models \bot$
A model $M = \langle W, R, S, \vdash \rangle$, 
$\vdash \subseteq W \times \text{Prop}$

- $w \not\vdash \bot$
- $w \vdash A \rightarrow B$ iff $w \not\vdash A$ or $w \vdash B$
A model $M = \langle W, R, S, \vdash \rangle$, 
$\vdash \subseteq W \times \text{Prop}$

- $w \not\vdash \bot$
- $w \vdash A \rightarrow B$ iff $w \not\vdash A$ or $w \vdash B$
- $w \vdash \square A$ iff $\forall v \ (wRv \Rightarrow v \vdash A)$
A model $M = \langle W, R, S, \models \rangle$, 
$\models \subseteq W \times \text{Prop}$

- $w \not\models \bot$
- $w \models A \rightarrow B$ iff $w \not\models A$ or $w \models B$
- $w \models \Box A$ iff $\forall v \ (wRv \Rightarrow v \models A)$
- $w \models A \triangleright B$ iff $\forall u \ (wRu \land u \models A \Rightarrow \exists v (uS_w v \models B))$
Montagna has a nice frame condition

\[(A \triangleright B) \rightarrow ((A \land \Box C) \triangleright (B \land \Box C))\]
Montagna has a nice frame condition

\[(A \triangleright B) \rightarrow ((A \land \Box C) \triangleright (B \land \Box C))\]

Beklemishev is somewhat similar
A B-simulation on a frame is a binary relation $S$ for which the following holds.

1. $S(x, x') \rightarrow x' = x'\uparrow$

2. $S(x, x') \& xRy \rightarrow \exists y'(ySy' \land S(y, y') \land y'S_{x'} \subseteq y_{x'} \uparrow) \land \forall d, e (y'S_{x}dRe \rightarrow yRd))$
\[
\begin{align*}
\text{ES}^0_2 &:= \text{ED}_2 \\
\text{ES}^{n+1}_2 &:= \text{ES}_2^n \mid \text{ES}^{n+1}_2 \land \text{ES}^{n+1}_2 \mid \text{ES}^{n+1}_2 \lor \text{ES}^{n+1}_2 \mid \\
&\quad \neg(\text{ES}_2^n \triangleright \text{Form})
\end{align*}
\]
\[
\begin{align*}
\text{ES}_2^0 & := \text{ED}_2 \\
\text{ES}_2^{n+1} & := \text{ES}_2^n \mid \text{ES}_2^{n+1} \land \text{ES}_2^{n+1} \mid \text{ES}_2^{n+1} \lor \text{ES}_2^{n+1} \mid \\
& \quad \neg (\text{ES}_2^n \upharpoonright \text{Form}) \\
S_0(b, u) & := b \uparrow = u \uparrow \\
S_{n+1}(b, u) & := S_n(b, u) \land \\
& \quad \forall c \ (bRc \rightarrow \exists c' \ (uRc' \land S_n(c, c') \land \\
& \qquad cS_bc' \land c'S_u \uparrow \subseteq cS_b \uparrow))
\end{align*}
\]
\[ \text{ES}_2^0 := \text{ED}_2 \]
\[ \text{ES}_2^{n+1} := \text{ES}_2^n \mid \text{ES}_2^{n+1} \land \text{ES}_2^{n+1} \lor \neg (\text{ES}_2^n \triangleright \text{Form}) \]
\[ S_0(b, u) := b \uparrow = u \uparrow \]
\[ S_{n+1}(b, u) := S_n(b, u) \land \forall c \ (b Rc \rightarrow \exists c' \ (u Rc' \land S_n(c, c') \land cS_b c' \land c' S_u \cup \subseteq cS_b \uparrow)) \]

For every \( i \) we define the frame condition \( C_i \) to be
\[ \forall a, b \ (a R b \rightarrow \exists u \ (b S_a u \land S_i(b, u) \land \forall d, e \ (u S_a d R e \rightarrow b R e)) \).
Why and how study interpretability
Proof theoretic characteristics of PRA
Modal matters

Frame conditions

\[ \text{ES}^0_2 := \text{ED}_2 \]

\[ \text{ES}^{n+1}_2 := \text{ES}^n_2 \land \text{ES}^{n+1}_2 \land \text{ES}^{n+1}_2 \lor \text{ES}^{n+1}_2 \land \neg (\text{ES}^n_2 \triangleright \text{Form}) \]

\[ S_0(b, u) := b \uparrow = u \uparrow \]

\[ S_{n+1}(b, u) := S_n(b, u) \land \forall c \ (bRc \rightarrow  \exists c' (uRc' \land S_n(c, c') \land cS_b c' \land c'S_u \uparrow \subseteq cS_b \uparrow)) \]

For every \( i \) we define the frame condition \( C_i \) to be

\[ \forall a, b \ (aRb \rightarrow  \exists u \ (bS_a u \land S_i(b, u) \land \forall d, e \ (uS_a dRe \rightarrow bRe))) \]
Why and how study interpretability
Proof theoretic characteristics of PRA
Modal matters
The basics
Frame conditions

\[ \begin{align*}
\text{ES}^0_2 & := \text{ED}_2 \\
\text{ES}^{n+1}_2 & := \text{ES}_2^n \land \text{ES}^{n+1}_2 \land \text{ES}^{n+1}_2 \lor \text{ES}^{n+1}_2 \lor \neg(\text{ES}_2^n \triangleright \text{Form}) \\
S_0(b, u) & := b \uparrow = u \uparrow \\
S_{n+1}(b, u) & := S_n(b, u) \land \\
& \quad \forall c \ (bRc \rightarrow \exists c' \ (uRc' \land S_n(c, c') \land \\
& \quad \quad cS_b c' \land c' S_u \uparrow \subseteq cS_b \uparrow)) \\
\end{align*} \]

For every \( i \) we define the frame condition \( C_i \) to be
\[ \forall a, b \ (aRb \rightarrow \exists u \ (bS_a u \land S_i(b, u) \land \forall d, e \ (uS_a dRe \rightarrow bRe))). \]

Theorem
A finite frame \( F \) validates all instances of Beklemishev’s principle if and only if \( \forall i \ F \models C_i \).
B \vdash Z
Why and how study interpretability
Proof theoretic characteristics of PRA
Modal matters

Frame condition Zambella?

Marta Bilkova†, Dick de Jongh*, and Joost J. Joosten*.
Interpretability in PRA
▶ $B \vdash Z$

▶ $B \models Z$

▶ Frame condition Zambella?