# Complexity of interpretability logics **IL**W, **IL**P and **IL**M

(work in progress)

Luka Mikec

University of Zagreb, University of Barcelona

Barcelona, 2019

# Interpretability

- Let $T_1$ and $T_2$ be some first order theories.
- Roughly, an interpretation of $T_2$ in $T_1$ is a pair $(f, U)$ where:
  - $f$ maps relational symbols to formulas;
  - $f(A \to B) = f(A) \to f(B)$ etc.;
  - $f(\forall x F) = \forall x (U(x) \to f(F))$ etc.;
  - for all sentences $F \in \mathcal{L}(T_2)$:

$$T_2 \vdash F \;\Rightarrow\; T_1 \vdash f(F).$$

- $T_1 \rhd T_2$: $T_1$ interprets $T_2$.

# Interpretability

- In particular, interpretability between finite extensions of a given theory:
$$T + A \rhd T + B$$
- Formalised interpretability: what properties of $\rhd$ can be proven in the base theory?

# Interpretability logics

- The language of interpretability logics is given by

$$A ::= p \mid \bot \mid A \to A \mid \Box A \mid A \rhd A,$$

  where $p$ is a propositional variable.
- Let $T$ be a formal theory, and $Int(\ulcorner A \urcorner, \ulcorner B \urcorner)$ a sentence formalizing $T + A \rhd T + B$.
- Arithmetical interpretation $*$ assigns sentences to modal formulas, such that:
  - $p^*$ is a sentence;
  - $(A \to B)^* = A^* \to B^*$ etc.;
  - $(\Box A)^* = Pr_T(A^*)$;
  - $(A \rhd B)^* = Int_T(A^*, B^*)$.

# Interpretability logics

- Given a theory $T$ (able to formalise interpretability),

$$A \in IL(T) \ :\Leftrightarrow \ \forall * \ T \vdash A^*.$$

- Interpretability logics of all “reasonable” theories contains the basic interpretability logic **IL**.

# Basic interpretability logic **IL**

- Basic interpretability logic **IL**:
  - propositionally valid formulas (in the new language);
  - K $\Box(A \to B) \to (\Box A \to \Box B)$;
  - Löb $\Box(\Box A \to A) \to \Box A$;
  - 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 $\Diamond A \rhd A$.
  - rules: modus ponens and necessitation $A/\Box A$.

  (parentheses priority: $\neg, \Box, \Diamond; \quad \wedge, \vee; \quad \rhd; \quad \to, \leftrightarrow$)
- $\Box A$ is **IL**-equivalent to $\neg A \rhd \bot$ (similarly for $\Diamond A$).

# Some principles

$$\mathbf{IL}\mathrm{P} = \mathbf{IL} + A \rhd B \to \Box(A \rhd B)$$
$$\mathbf{IL}\mathrm{M} = \mathbf{IL} + A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C$$
$$\mathbf{IL}\mathrm{W} = \mathbf{IL} + A \rhd B \to A \rhd (B \wedge \Box\neg A)$$

- $IL(T) = \mathbf{IL}\mathrm{P}$ iff $T$ is finitely axiomatizable and sufficiently strong;
- $IL(T) = \mathbf{IL}\mathrm{M}$ iff $T$ proves full induction;
- $IL(T) \supsetneq \mathbf{IL}\mathrm{W}$.

# Models

- Semantics: extend the usual relational (Kripke) model.
- **IL**-frame (Veltman frame): $\mathcal{F} = \langle W, R, \underline{\{S_w : w \in W\}} \rangle$, where:
  1. $W \neq \emptyset$;
  2. $R^{-1}$ is well-founded (no $x_0 R x_1 R x_2 R \ldots$ chains);
  3. $R$ is transitive;
  4. $S_w \subseteq R[w]^2$ is reflexive, transitive, contains $R \cap R[w]^2$ ($wRuRv$ implies $uS_wv$);
- **IL**-model (Veltman model): $\mathcal{M} = \langle W, R, \{S_w : w \in W\}, V \rangle$, where:
  1. $\langle W, R, \{S_w : w \in W\} \rangle$ is an **IL**-frame;
  2. $V \subseteq W \times Prop$ (or $V : Prop \to \mathcal{P}(W)$).

# Models

- Veltman model: $\mathcal{M} = \langle W, R, \underline{\{S_w \ : \ w \in W\}}, V \rangle$.
- $w \Vdash p$ if and only if $wVp$, for $p \in Prop$.
- Logical connectives have classical semantics.
- Truth of a formula $F \rhd G$ ("$F$ interprets $G$") in a world $w \in \mathcal{M}$:

$$w \Vdash F \rhd G \quad :\Leftrightarrow \quad \forall x \in R[w] : \quad x \Vdash F \Rightarrow \exists y \in S_w(x) : \ y \Vdash G.$$

- Modal soundness and completeness:

$$IL \vdash F \quad \Leftrightarrow \quad \forall \mathcal{F} : \mathcal{F} \vDash F.$$

# Extensions and frame conditions

$$
\begin{array}{ll}
ILP & IL + A \rhd B \to \Box(A \rhd B) \\
ILM & IL + A \rhd B \to A \wedge \Box C \rhd B \wedge \Box C \\
ILW & IL + A \rhd B \to A \rhd B \wedge \Box \neg A
\end{array}
$$

- These logics are complete w.r.t. certain classes of frames:

$$
\begin{array}{ll}
(P) & wRw'RuS_w v \Rightarrow uS_{w'} v; \\
(M) & wRuS_w v \Rightarrow R[v] \subseteq R[u]; \\
(W) & S_w \circ R \text{ is converse well-founded for each } w;
\end{array}
$$

- **IL**W-frame is **IL**-frame that satisfies $(W)$ etc.

# Complexity

- **IL** conservatively extends **GL** ("provability logic"); **GL** is in $\mathtt{PSPACE}$.
- Closed fragment of **IL** is $\mathtt{PSPACE}$-hard (Bou, Joosten).
- FMP for **IL**: if $x \Vdash F$, then there is a finite $\mathcal{M}$ and $x' \in \mathcal{M}$ s.t. $x' \Vdash F$.
- Standard approach: to check if $\vdash F$, we can (soundness, completeness, FMP) check if there is a finite model of $\neg F$.
- So, to prove $IL \in \mathtt{PSPACE}$, it suffices to construct a $\mathtt{PSPACE}$ algorithm that tests satisfiability.

# Complexity (satisfiability)

- A natural approach would be to build the model one world at a time.
- If $A \triangleright B \notin w$, try modelling a $B$-critical world satisfying $A$.
- If $A \triangleright B \in w, A \in x$, try modelling $B$ with the same criticality as $x$.
- A very naive implementation does not terminate.
- But similarly with less naive approaches that we tried.

# Complexity of **IL**

- Let $\Gamma$ be an adequate set for $A \in \mathcal{L}$: a set of subformulas closed under certain operations.
- $|\Gamma|$ is polynomial in $|A|$.
- Our algorithm builds models piece-by-piece (nondeterministically or with backtracking), where each "piece" is a (small) set of worlds.
- We introduce functions named (1), (2) and (3).
- (1) only calls (2), which only calls (3), which only calls (1).

# Function (1)

- (1) takes $\Delta \subseteq \Gamma$ and checks whether there is a rooted Veltman model of $\Delta$ ($W = \{w\} \cup R[w]$, $w \Vdash \Delta$)
- The starting call will be with $\Delta = \{A\}$.
- (1) looks at all the maximal Boolean consistent $\Delta' \supseteq \Delta$, and returns a positive result if at least one extension is satisfiable.
- Lemma: (1) returns a positive result if and only if $\Delta$ is satisfiable.

# Function (2)

- (2) takes a maximal Boolean consistent $\Delta \subseteq \Gamma$ and checks whether there is a rooted Veltman model of $\Delta$.

$$\Delta^+ := \{A \rhd B \in \Gamma : A \rhd B \in \Delta\}$$
$$\Delta^- := \{A \rhd B \in \Gamma : \neg(A \rhd B) \in \Delta\}$$

- (2) returns a positive answer if the sets $\{\neg(C \rhd D)\} \cup \Delta^+$ are satisfiable for all $\neg(C \rhd D) \in \Delta^-$.
- Lemma: (2) returns a positive result if and only if $\Delta$ is satisfiable. (Proof: by merging roots)

# Function (3)

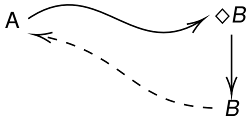
- (3) takes a Boolean consistent $\Delta \subseteq \Gamma$ consisting of one negated $\rhd$-formula $\neg(C \rhd D)$ and a set of positive $\rhd$-formulas $\Delta^+$, and checks whether there is a model of $\Delta$.
- We say that $(N, P)$ is a $(\neg(C \rhd D), \Delta)$-pair if:
  1. $N, P \subseteq \Gamma$;
  2. $D \in N$;
  3. $\bot \notin P$;
  4. $A \rhd B \in \Delta^+ \;\Rightarrow\; A \in N$ or $B \in P$.
- (3) returns a positive answer if there is a $(\neg(C \rhd D), \Delta)$-pair $(N, P)$ such that the following holds:
  1. $\{\neg A, A \rhd \bot \mid A \in N\} \cup \{C, C \rhd \bot\}$ is satisfiable;
  2. $\{\neg A, A \rhd \bot \mid A \in N\} \cup \{B, B \rhd \bot\}$ is satisfiable for all $B$ in $P$.
- Lemma: (3) returns a positive result if and only if $\Delta$ is satisfiable. (Proof: by joining the models, adding a new root $w$, and extending $S_w$ where needed – or even making it total).

# Wrapping up (**IL**)

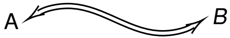
- Note that (1) can be calculated in terms of (2) etc.
- Each (1)-(2)-(3) chain adds a new $\Box\neg B$ formula for some $B \in \Gamma$; the procedure terminates.
- Algorithm works locally correct: each function does what it is supposed to do assuming the next one does. Full correctness by induction (starting with leaf nodes in the execution tree).
- **IL** was known to be $\mathtt{PSPACE}$-hard (conservatively extends **GL**; also $\mathbf{IL}_0$). Thus, **IL** is $\mathtt{PSPACE}$-complete.

# **IL**W

- Preventing $(R \circ S_w)$-loops.
- Assume we have $\Diamond A$, $A \rhd \Diamond B \vee \Diamond C$ and $B \rhd A$. Our algorithm for **IL** might build:



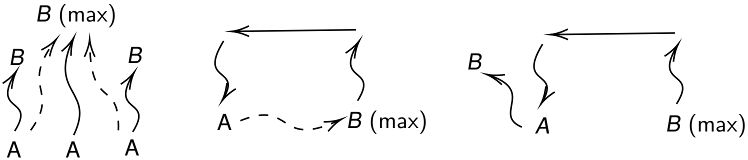

- But not all $S_w$-loops are bad. Assume $\Diamond A$, $A \rhd B$ and $B \rhd A$.




- We also can't make $S_w$ total as before.

# **IL**W

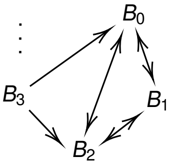
- Solution: ensure each witness is $(S_w \circ R \circ S_w)$-maximal.
- Lemma: in any cone witnessing $\neg(C \rhd D)$ and $A_i \rhd B_i$, we can $S_w$-connect $A_i$ to $(S_w \circ R \circ S_w)$-maximal witnesses of $B_i$.




- $(S_w \circ R \circ S_w)$-maximality is lost in the process, but this can be fixed.

# **IL**W

- Algorithm: iterate through ($S_w$-) “visibility” graphs in advance.



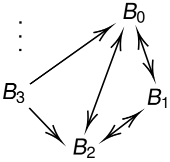

- Existence of an arrow $A \to B$: the witness of $A$ can $S_w$-see the witness of $B$;
- Non-existence: the witness of $A$ can’t $S_w$-see *any* $B$ (Lemma ensures that this is wlog.)

# **IL**W

- Visibility graphs are of polynomial size ($\sim |\Gamma|^2$).
- Reflexive and transitive, like $S_w$.
- Two kinds of information:
  1. if $A \not\twoheadrightarrow B$, submodel generated by the witness of $A$ should not entail (anything that triggers) $B$.
  2. if $A \longleftrightarrow B$: everything in cluster should forbid everything in cluster after an $R$-transition.

# **IL**W

- Previous example:




- We have a cluster $\mathcal{C} = \{B_0, B_1, B_2\}$.
- Since $B_0 \longrightarrow\!\!\!\!\!/\;\; B_3$, (the witness for) $B_0$ can't $S_w$-see anything that triggers $B_3$.
- If $B_i, B_j \in \mathcal{C}$ and $E \rhd B_j$, (the witness for) $B_0$ can't $S_w$-see $E$.

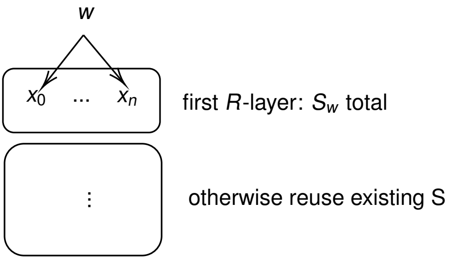
# **IL**P

- Whenever $uS_wv$, also $uS_{w'}v$, for any $w'$ between $w$ and $u$.
- Assume we have a cone witnessing $\neg(C \rhd D)$, $A_i \rhd B_i$.
- Approach: when prepending $w$ to worlds $x_i$ witnessing $\neg(C \rhd D)$, $A_i \rhd B_i$,

$$\{A \rhd B : A \rhd B \in w\} \subseteq \bigcap \{A \rhd B : A \rhd B \in x_i\}$$

# ILP

- Immediate successors of $w$ should be $S_w$-connected.
- If $A \rhd B \in w$, $wRx \Vdash A$ and $x$ is not an immediate successor of $w$:
  Let $x_i$ be the immediate successor of $w$, a predecessor of $x$.
  Then $xS_{x_i}y \Vdash B$. Define $xS_wy$.

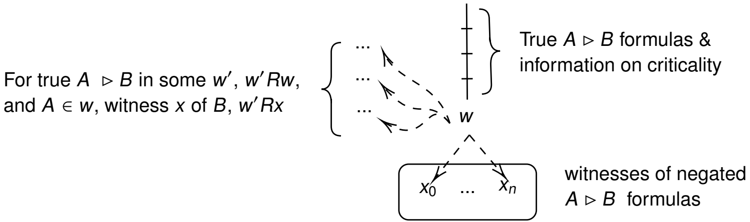


# **IL**M

- Whenever $wRuS_wvRz$, ensure $uRz$. Essentially $u \subseteq_\Box v$.
- For **IL**, **IL**W, and (in some sense) **IL**P, each piece of model required a polynomial (in fact linear) amount of worlds.
- Can this be done with **IL**M?

# **IL**M

- Instead, here we can use the "naive" approach.




- Principle M is just strong enough to make this viable:
  - With $\neg(A \rhd B) \in w'$, try obtaining $A$ in a $B$-critical cone.
  - With $A \rhd B \in w'Rw$ and $A \in w$, either reuse an *available witness* $x$ (if any), or create a new world.
- At most $n = |\Gamma|$ boxed formulas. At most $n$ calls resulting in reusable worlds. At most $n$ level decreases. So, the maximal depth of a call tree is $n^3$.

Thank you.

Implementations (to be updated):
`https://github.com/luka-mikec/provability_sat`

Previous work:


- L. Mikec, F. Pakhomov, M. Vuković. Complexity of the interpretability logic **IL**. Logic Journal of the IGPL, 2018.



This work has been supported by the Croatian Science Foundation, grants UIP-05-2017-9219 and IP-01-2018-7459.