Equilibrium models for epistemic specifications

Luis Fariñas del Cerro, Andreas Herzig, Ezgi Iraz Su

University of Toulouse and CNRS, IRIT, France

A Coruña, Feb. 24, 2015
From logic to logic programs

- logic: $\varphi \rightarrow \psi$
- logic programming: $Head \leftarrow Body$
  - $Head$ disjunction of atoms
  - $Body$ conjunction of atoms, possibly prefixed by “not”
    - ‘default negation’, ‘negation by failure’ = non-deducibility of $p$
    - no consensus on semantics until the 90ies
  - disregarded here: second, 3-valued (‘strong’) negation “$\overline{p}$”
    (compiled away: replace $\overline{p}$ by new variable $p'$ and add $\leftarrow p, p'$)
- answer set semantics
  - fixed point definition: $I$ is an answer set for $\Pi$ iff $I = reduct(\Pi, I)$
  - remarkably ‘stable’: there exist 10+ different characterisations
    [Lifschitz “Twelve Definitions of a Stable Model”, ICLP 2008]
Towards a logical account of negation by failure

- hypothesis: not every classical model of a program intended (identifying not with ¬)
- models should minimize truth of atoms
  - example: \( \Pi = p \leftarrow p \) has unique minimal model \( \emptyset \)
  - so every \( p \) is false
- problem: programs such as \( \{ p \leftarrow \neg p \} \) should have no model
  - ...but \( \neg p \rightarrow p \) is equivalent to \( p \) in classical logic
- solution: \( \neg p \rightarrow p \) is not equivalent to \( p \) in intuitionistic logic
  (more generally: intermediate logics)
The logic of here-and-there (HT)

- simple modal logic:
  - only two possible worlds $H$ (‘here’) and $T$ (‘there’)
  - accessibility relation is reflexive, and $T$ is accessible from $H$
  - idea: $H = \text{proved true, } T = \text{hypothesised, } \text{PVAR} \setminus T = \text{refuted}$

- is an intuitionistic logic:
  - $H \subseteq T$ (‘heredity condition’)
  - interprets a language with a connective $\rightarrow$ that is stronger than material implication $\supset$
    - $\models \neg \varphi \leftrightarrow (\varphi \rightarrow \bot)$
    - $\models \varphi \rightarrow \neg \neg \varphi$
    - $\not\models \varphi \leftrightarrow \neg \neg \varphi$
    - $\not\models \varphi \lor \neg \varphi$
    - $\not\models \varphi \lor \neg \varphi$
The logic of here-and-there (HT)

- ht-model = \((H, T)\) such that \(H \subseteq T \subseteq \text{PVAR}\)
  - \(H = T\): ‘total model’
- truth conditions:
  
  \[
  \begin{align*}
  H, T \models p & \text{ iff } p \in H \\
  H, T \models \neg \varphi & \text{ iff } T, T \not\models \varphi \\
  H, T \models \varphi \rightarrow \psi & \text{ iff } H, T \models \varphi \supset \psi \text{ and } T, T \models \varphi \supset \psi
  \end{align*}
  \]
  (where \(\supset\) is material implication)

Theorem (Lifschitz et al. 2001)

\[\Pi_1 \text{ and } \Pi_2 \text{ are strongly equivalent} \iff \models_{\text{HT}} \Pi_1 \leftrightarrow \Pi_2\]

(Identifying not with \(\neg\))
Equilibrium models

- equilibrium model: $H = T$ (total model) such that there is no smaller $ht$-model

**Definition**

$\langle T, T \rangle$ equilibrium model of $\varphi$ iff

1. $T, T \models \varphi$
2. $H, T \not\models \varphi$ for every $H \subset T$

**Theorem (Pearce 1996)**

$\langle T, T \rangle$ equilibrium model of $\Pi$ iff $T$ answer set of $\Pi$

*(identifying “not” with “¬”)*

- applies beyond standard logic programs
  - disjunctive logic programs: $H = p \lor q$
  - nested logic programs: $B = p \leftarrow (q \leftarrow r)$
  - ...

where the 10+ semantics don’t agree!
- missing: quantification over possible answer sets...
ASP lacks expressivity

Example (scholarship eligibility program)

1. eligible ← highGPA
2. eligible ← minority, fairGPA
3. eligible ← fairGPA, highGPA
4. interview ← not eligible, not eligible
5. fairGPA or highGPA ←

has the answer sets

\[ \text{AS}(\Pi_{\text{eligible}}) = \left\{ \{\text{highGPA, eligible}\}, \right\} \{\text{fairGPA}\} \right\} \]

Therefore:

\[ \Pi_{\text{eligible}} \not\models \text{eligible} \]
\[ \Pi_{\text{eligible}} \not\models \text{interview} \]

⇒ counter-intuitive!
ASP lacks expressivity

Example (scholarship eligibility program)

1. eligible ← highGPA
2. eligible ← minority, fairGPA
3. eligible ← fairGPA, highGPA
4. interview ← not eligible, not eligible
5. fairGPA or highGPA ←

has the answer sets

\[ AS(\Pi_{\text{eligible}}) = \left\{ \{\text{highGPA, eligible}\}, \{\text{fairGPA}\} \right\} \]

Therefore:

\[ \Pi_{\text{eligible}} \not\models \text{eligible} \]
\[ \Pi_{\text{eligible}} \not\models \text{interview} \]

⇒ counter-intuitive!
Epistemic specifications [Gelfond 1991]

Example (scholarship eligibility program, E-S-version)

1. \( \text{eligible} \leftarrow \text{highGPA} \)
2. \( \text{eligible} \leftarrow \text{minority}, \text{fairGPA} \)
3. \( \text{eligible} \leftarrow \text{fairGPA}, \text{highGPA} \)
4. \( \text{interview} \leftarrow \neg K \text{eligible}, \neg K \text{eligible} \)
5. \( \text{fairGPA or highGPA} \leftarrow \)

will have the answer sets

\[
\text{AS}(\Pi_{\text{K eligible}}) = \{ \{\text{highGPA, eligible, interview}\}, \\
\{\text{fairGPA, interview}\} \}
\]

Therefore:

\( \Pi_{\text{K eligible}} \not\models \text{eligible} \)
\( \Pi_{\text{K eligible}} \models \text{interview} \)
### Epistemic specifications [Gelfond 1991]

**Example (scholarship eligibility program, E-S-version)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>eligible ← highGPA</td>
</tr>
<tr>
<td>2</td>
<td>eligible ← minority, fairGPA</td>
</tr>
<tr>
<td>3</td>
<td>eligible ← fairGPA, highGPA</td>
</tr>
<tr>
<td>4</td>
<td>interview ← not K eligible, not K eligible</td>
</tr>
<tr>
<td>5</td>
<td>fairGPA or highGPA ←</td>
</tr>
</tbody>
</table>

will have the answer sets

\[
\text{AS}(\Pi_{\text{eligible}}) = \left\{ \left\{ \text{highGPA, eligible, interview} \right\}, \left\{ \text{fairGPA, interview} \right\} \right\}
\]

Therefore:

\[
\Pi_{\text{eligible}} \not\approx \text{eligible} \\
\Pi_{\text{eligible}} \approx \text{interview}
\]
Epistemic specifications: language

- Idea: allow for quantification over all candidate answer sets
  - $K q = “it is known that q”$
  - $M q = “q may be believed”$
    (more standard: “compatible with the agent’s knowledge”)

- Syntax of rules varies from paper to paper, but basically interdefinable

- Grammar [Kahl 2014]:

  \[ l_1 \text{ or } \ldots \text{ or } l_k \leftarrow \lambda_1, \ldots, \lambda_m \]

  - Head: objective literals $l, l_1, l_2, \ldots$ (possibly strongly negated)
  - Body: extended literals

  \[ \lambda ::= l \mid \text{not} \ l \mid 
  
  K \ l \mid \text{not} K \ l \mid 
  M \ l \mid \text{not} M \ l \]
Epistemic specifications: semantics

- idea:
  1. move from answer sets to world views = sets of answer sets
  2. reduct $\Pi^W$ of an epistemic specification $\Pi$ by a world view $W$ (eliminates modal operators)
     \[ \Rightarrow \text{procedural} \]
  3. fixpoint equation defines sets of answer sets
     \[ \Rightarrow \text{non-constructive} \]

- still no consensus on reduct definition
  - [Gelfond, AMAI 1994]
  - [Gelfond, LPNMR 2011]
  - [Kahl, PhD 2014]

- ht-logic and equilibrium logic counterpart?
  - [Wang&Zhang, LPNMR 2005], v.i.
  - [FHS], v.i.
Epistemic specifications: semantics

idea:
1. move from answer sets to world views = sets of answer sets
2. reduct $\Pi^W$ of an epistemic specification $\Pi$ by a world view $W$ (eliminates modal operators) $\Rightarrow$ procedural
3. fixpoint equation defines sets of answer sets $\Rightarrow$ non-constructive

still no consensus on reduct definition
- [Gelfond, AMAI 1994]
- [Gelfond, LPNMR 2011]
- [Kahl, PhD 2014]

ht-logic and equilibrium logic counterpart?
- [Wang&Zhang, LPNMR 2005], v.i.
- [FHS], v.i.
**Definition**

- **reduct** $\Pi^W$ of an epistemic specification $\Pi$ by a world view $W$: for each rule,

<table>
<thead>
<tr>
<th>literal in body:</th>
<th>if true in $W$:</th>
<th>if false in $W$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K/l$</td>
<td>replace by $l$</td>
<td>delete rule</td>
</tr>
<tr>
<td>$\neg K/l$</td>
<td>replace by $\top$</td>
<td>replace by $\neg l$</td>
</tr>
<tr>
<td>$M/l$</td>
<td>replace by $\top$</td>
<td>replace by $\neg \neg l$</td>
</tr>
<tr>
<td>$\neg M/l$</td>
<td>replace by $\neg l$</td>
<td>delete rule</td>
</tr>
</tbody>
</table>
Problem 1: cycle with $K$

\[ \Pi_{18} = \{ p \leftarrow Kp \} \]


- has unique world view $\{\emptyset\}$ [Gelfond 2011, Kahl 2014, FHS]

Remark. clear case: $Kp \rightarrow p$ is the truth axiom of epistemic logic
Problem 2: cycle with $\mathcal{M}$

\[ \Pi_1 = \{ p \leftarrow \mathcal{M} p \} \]

- has unique world view \( \{ \{ p \} \} \) \[\text{[Kahl 2014]}\]
- has 2 world views \( \{ \emptyset \} \) and \( \{ \{ p \} \} \) \[\text{[Gelfond 1991, 1994], [Wang & Zhang 2005]}\]
- has unique world view \( \{ \emptyset \} \) \[\text{[FHS]}\]
  - has 2 world views \( \{ \emptyset \} \) and \( \{ \{ p \} \} \) if $\mathcal{M}$ replaced by \( \neg K \neg \) \[\text{[FHS]}\]

Remark. circular $\Rightarrow$ no clear intuitions (at least for us)
Problem 3: preference over a disjunction

\[ \Pi_{32} = \{ p \text{ or } q \leftarrow, \ q \leftarrow \text{ M } p \} \]

- has unique world view \( \{ q \} \) [Kahl 2014, FHS]

Remark. intuitively clear (similar to Gelfond’s eligibility example)
Problem 4: preference over a disjunction, ctd.

\[ \Pi_{32} = \{ p \text{ or } q \leftarrow, \ q \leftarrow \text{not } K\ p \} \]

- ☺ has unique world view \( \{ q \} \)  
  [Kahl 2014]

- ☹ has 2 world views \( \{ q \} \) and \( \{ p \} \)  

Remark. intuitively clear (similar to Gelfond’s eligibility example)
[Wang & Zhang 2005]'s epistemic extension of HT

- ‘occamist’ combination of ht-models and K45
- WZ-model = (W, H, T) where
  - W is a classical S5 model: W ⊆ 2^{PVAR}
  - (H, T) is an ht-model: H ⊆ T ⊆ PVAR
    ⇒ H and T not necessarily in W (!)

- truth conditions:
  - W, H, T ⊨ Kφ iff W, H’, T’ ⊨ φ for every ht-model H’, T’ that can be built from W

- <W, T, T> is an epistemic equilibrium model of φ iff
  - <W, T, T> ⊨ φ and <W, H, T> /∈ φ for every H ⊂ T

- <W> is an equilibrium view of φ iff W is the maximal collection satisfying W = {T : <W, T, T> is an epi.eq.model of φ}

Theorem (Wang&Zhang 2005, Thm. 2)

W is a world view of Π iff W is an equilibrium view of Π.
[Wang & Zhang 2005]’s epistemic extension of HT: criticisms

1. not really an epistemic logic
   - $p \land K \neg p$ has a model (and even a WZ-equilibrium model)

2. not really an intuitionistic modal logic
   - $K \varphi \leftrightarrow \neg M \neg \varphi$ valid
   - $K \neg \neg \varphi \rightarrow K \varphi$ valid
   - $\neg \neg K \varphi \rightarrow K \varphi$ valid

3. equilibrium definition unintuitive beyond disjunctive logic programs (‘nested epistemic logic programs’, NELP)
   - $(\mathcal{W}, T, T)$ is WZ-equilibrium model of $K p$ iff $\mathcal{W}$ S5-model of $K p$ and $T=\emptyset$
     - no minimisation
   - $K p$ has no WZ-equilibrium model
   - $M p \land M \neg p$ has no WZ-equilibrium view
Our approach

1. standard epistemic extension of HT
two-dimensional modal logic (cf. intuitionistic S5)

2. maximise falsehood: cf. equilibrium logic
   - $\emptyset \not\models_{EE} K \neg p$
   - $p \lor q \not\models_{EE} K (p \lor q)$
   - $p \lor q \not\models_{EE} M p \land M q$

3. maximise ignorance: cf. Levesque’s “all-that-I-know” and Moore’s autoepistemic logic
   - $p \lor q \models_{AEE} M p \land M q$
   - however makes no difference for the discriminating examples
Our epistemic ht-models

two-dimensional modal logic (cf. intuitionistic S5)

Definition

e-ht-model = (\mathcal{W}, \bar{h}) where

- \mathcal{W} is a classical S5 model: \mathcal{W} \subseteq 2^{P\text{VAR}}
- \bar{h} : \mathcal{W} \rightarrow 2^{P\text{VAR}} such that \bar{h}(T) \subseteq T for every \ T \in \mathcal{W}

- classical S5 model: \bar{h} = \text{id}
- truth conditions:
  \begin{align*}
  (\mathcal{W}, \bar{h}), T \models p & \iff p \in \bar{h}(T) \\
  (\mathcal{W}, \bar{h}), T \models \varphi \rightarrow \psi & \iff (\mathcal{W}, \bar{h}), T \models \varphi \supset \psi \text{ and } (\mathcal{W}, \text{id}), T \models \varphi \supset \psi \\
  (\mathcal{W}, \bar{h}), T \models \text{K}\varphi & \iff (\mathcal{W}, \bar{h}), T' \models \varphi \text{ for every } T' \in \mathcal{W} \\
  (\mathcal{W}, \bar{h}), T \models \text{M}\varphi & \iff (\mathcal{W}, \bar{h}), T' \models \varphi \text{ for some } T' \in \mathcal{W}
  \end{align*}

- satisfies the requirements for intuitionistic modal logics

Our epistemic equilibrium models

minimise truth (cf. equilibrium logic)

**Definition**

\( \mathcal{W} \) is an epistemic equilibrium model of \( \varphi \) iff

1. \((\mathcal{W}, \text{id}), T \models \varphi \) for every \( T \in \mathcal{W} \) (classical S5 model of \( \varphi \))

2. There is no \( \mathcal{h} \neq \text{id} \) such that \((\mathcal{W}, \mathcal{h}), T \models \varphi \) for every \( T \in \mathcal{W} \) (no ‘weaker’ e-ht-model of \( \varphi \))

**Example:** \{\( p \) or \( \overline{p} \) \} has 3 epistemic eq. models:

- \( \emptyset \), \( \{p\} \), and \( \emptyset, \{p\} \)

**Theorem (strong equivalence)**

...
Our autoepistemic equilibrium models

minimise knowledge (cf. Levesque’s “all-that-I-know”)

**Definition**

\((\mathcal{W}, T)\) is an autoepistemic equilibrium model of \(\varphi\) iff

1. \((\mathcal{W}, T)\) is an epistemic equilibrium model of \(\varphi\)
2. \((\mathcal{W}', T)\) is not an epistemic equilibrium model of \(\varphi\), for every \(\mathcal{W}'\) such that \(\mathcal{W}' \supseteq \mathcal{W}\) (no ‘bigger’ epi.eq.model of \(\varphi\))

**Example:** \(\{p \text{ or } \bar{p} \leftarrow\}\) has 1 autoepistemic eq.model:

\[\{\emptyset, \{p\}\}\]

**Theorem (strong equivalence)**

...
Ongoing work: first minimise knowledge, then truth?

- given $\Pi$, 
  1. compute the biggest S5 model $W$ of $\Pi$
  2. compute the biggest subset of $W$ that is an epistemic eq.model

- gets right all the examples but $p \leftarrow M p$
To sum it up

- problem with preference over disjunctions: [Gelfond 2005]
- gets all examples right (idea of support): [Kahl 2014]
- epistemic HT good basis for further work:
  - simple intuitionistic modal logic
  - epistemic equilibrium models (minimises truth)
  - autoepistemic equilibrium models (maximises ignorance)
- programs with cycles:
  - intuitions not clear (perhaps not only for us)
  - semantics not easy to define