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: $\varphi \to \psi$
- 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 Π iff $I = reduct(\Pi, I)$
  - remarkably 'stable': there exist 10+ different characterisations

    [Lifschitz "Twelve Definfitions 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 $\neg$)
- 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 $\{p \leftarrow \texttt{not}\, p\}$ 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 = proved true, T = hypothesised, PVAR $\setminus T$ = refuted
- is an intuitionistic logic:
  - $H \subseteq T$ ('heredity condition')
  - interprets a language with a connective $\rightarrow$ that is stronger than material implication $\supset$

$$\models \neg\varphi \leftrightarrow (\varphi \rightarrow \bot)$$
$$\models \varphi \rightarrow \neg\neg\varphi$$
$$\not\models \varphi \leftarrow \neg\neg\varphi$$
$$\not\models \varphi \vee \neg\varphi$$

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:

$$\begin{aligned} H, T &\models p \;\text{ iff } p \in H \\ H, T &\models \neg\varphi \;\text{ iff } T, T \not\models \varphi \\ H, T &\models \varphi \rightarrow \psi \;\text{ iff } H, T \models \varphi \supset \psi \text{ and } T, T \models \varphi \supset \psi \end{aligned}$$

(where $\supset$ is material implication)

Theorem (Lifschitz et al. 2001)

Π_1 *and* Π_2 *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

(T, T) equilibrium model of φ iff

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

Theorem (Pearce 1996)

(T, T) equilibrium model of Π iff T answer set of Π

(identifying "`not`" with "$\neg$")

- applies beyond standard logic programs
  - disjunctive logic programs: $H = p \;\texttt{or}\; q$
  - 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. $\mathtt{eligible} \leftarrow \mathtt{highGPA}$
2. $\mathtt{eligible} \leftarrow \mathtt{minority},\ \mathtt{fairGPA}$
3. $\overline{\mathtt{eligible}} \leftarrow \overline{\mathtt{fairGPA}},\ \overline{\mathtt{highGPA}}$
4. $\mathtt{interview} \leftarrow \mathrm{not}\, \mathtt{eligible},\ \mathrm{not}\, \overline{\mathtt{eligible}}$
5. $\mathtt{fairGPA}\ \mathrm{or}\ \mathtt{highGPA} \leftarrow$

has the answer sets

$$AS(\Pi_{\mathtt{eligible}}) = \Big\{ \{\mathtt{highGPA},\ \mathtt{eligible}\},\ \{\mathtt{fairGPA}\} \Big\}$$

Therefore:

$$\Pi_{\mathtt{eligible}} \not\models \mathtt{eligible}$$

$$\Pi_{\mathtt{eligible}} \not\models \mathtt{interview}$$

$\Rightarrow$ counter-intuitive!

ASP lacks expressivity

Example (scholarship eligibility program)

1. $\texttt{eligible} \leftarrow \texttt{highGPA}$
2. $\texttt{eligible} \leftarrow \texttt{minority}, \texttt{fairGPA}$
3. $\overline{\texttt{eligible}} \leftarrow \overline{\texttt{fairGPA}}, \overline{\texttt{highGPA}}$
4. $\texttt{interview} \leftarrow \textbf{not}\, \texttt{eligible}, \textbf{not}\, \overline{\texttt{eligible}}$
5. $\texttt{fairGPA}\, \textbf{or}\, \texttt{highGPA} \leftarrow$

has the answer sets

$$AS(\Pi_{\texttt{eligible}}) = \Big\{ \{\texttt{highGPA}, \texttt{eligible}\}, \{\texttt{fairGPA}\} \Big\}$$

Therefore:

$$\Pi_{\texttt{eligible}} \not\models \texttt{eligible}$$

$$\Pi_{\texttt{eligible}} \not\models \texttt{interview}$$

⇒ counter-intuitive!

Epistemic specifications [Gelfond 1991]

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

1. $\texttt{eligible} \leftarrow \texttt{highGPA}$
2. $\texttt{eligible} \leftarrow \texttt{minority}, \texttt{fairGPA}$
3. $\overline{\texttt{eligible}} \leftarrow \overline{\texttt{fairGPA}}, \overline{\texttt{highGPA}}$
4. $\texttt{interview} \leftarrow \textbf{not}\, \mathbf{K}\, \texttt{eligible}, \textbf{not}\, \mathbf{K}\, \overline{\texttt{eligible}}$
5. $\texttt{fairGPA}\ \textbf{or}\ \texttt{highGPA} \leftarrow$

will have the answer sets

$$AS(\Pi_{\mathbf{K}\,\texttt{eligible}}) = \Big\{ \{\texttt{highGPA}, \texttt{eligible}, \texttt{interview}\}, \{\texttt{fairGPA}, \texttt{interview}\} \Big\}$$

Therefore:

$$\Pi_{\mathbf{K}\,\texttt{eligible}} \not\models \texttt{eligible}$$

$$\Pi_{\mathbf{K}\,\texttt{eligible}} \models \texttt{interview}$$

Epistemic specifications [Gelfond 1991]

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

1. $\mathtt{eligible} \leftarrow \mathtt{highGPA}$
2. $\mathtt{eligible} \leftarrow \mathtt{minority}, \ \mathtt{fairGPA}$
3. $\overline{\mathtt{eligible}} \leftarrow \overline{\mathtt{fairGPA}}, \ \overline{\mathtt{highGPA}}$
4. $\mathtt{interview} \leftarrow \mathtt{not}\, \mathbf{K}\, \mathtt{eligible}, \ \mathtt{not}\, \mathbf{K}\, \overline{\mathtt{eligible}}$
5. $\mathtt{fairGPA}\ \mathtt{or}\ \mathtt{highGPA} \leftarrow$

will have the answer sets

$$AS(\Pi_{\mathbf{K}\,\mathtt{eligible}}) = \Big\{ \{\mathtt{highGPA}, \ \mathtt{eligible}, \ \mathtt{interview}\}, \ \{\mathtt{fairGPA}, \ \mathtt{interview}\} \Big\}$$

Therefore:

$$\Pi_{\mathbf{K}\,\mathtt{eligible}} \not\approx \mathtt{eligible}$$

$$\Pi_{\mathbf{K}\,\mathtt{eligible}} \approx \mathtt{interview}$$

Epistemic specifications: language

- idea: allow for quantification over all candidate answer sets
  - $\mathtt{K}\,q$ = "it is known that q"
  - $\mathtt{M}\,q$ = "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]:

$$\mathtt{l}_1 \;\mathtt{or}\; \ldots \;\mathtt{or}\; \mathtt{l}_k \leftarrow \lambda_1, \ldots, \lambda_m$$

  - head: objective literals $\mathtt{l}, \mathtt{l}_1, \mathtt{l}_2, \ldots$ (possibly strongly negated)
  - body: extended literals

$$\lambda ::= \mathtt{l} \mid \mathtt{not}\,\mathtt{l} \mid \mathtt{K}\,\mathtt{l} \mid \mathtt{not}\,\mathtt{K}\,\mathtt{l} \mid \mathtt{M}\,\mathtt{l} \mid \mathtt{not}\,\mathtt{M}\,\mathtt{l}$$

Epistemic specifications: semantics

- idea:
  1. move from answer sets to world views = sets of answer sets
  2. reduct $\Pi^{\mathcal{W}}$ of an epistemic specification Π by a world view $\mathcal{W}$ (eliminates modal operators)

     $\Rightarrow$ procedural
  3. fixpoint equation defines sets of answer sets

     $\Rightarrow$ non-constructive
- still no consensus on reduct definition
  - [Gelfond, Tech.Rep. 1991]
  - [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 $\Pi^{\mathcal{W}}$ of an epistemic specification Π by a world view $\mathcal{W}$ (eliminates modal operators)

     $\Rightarrow$ procedural
  3. fixpoint equation defines sets of answer sets

     $\Rightarrow$ non-constructive
- still no consensus on reduct definition
  - [Gelfond, Tech.Rep. 1991]
  - [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: reducts [Kahl 2014]

Definition

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

literal in body:	if true in $\mathcal{W}$:	if false in $\mathcal{W}$:
$\mathtt{K}\,l$	replace by l	delete rule
$\mathtt{not}\,\mathtt{K}\,l$	replace by $\top$	replace by $\mathtt{not}\,l$
$\mathtt{M}\,l$	replace by $\top$	replace by $\mathtt{not}\,\mathtt{not}\,l$
$\mathtt{not}\,\mathtt{M}\,l$	replace by $\mathtt{not}\,l$	delete rule

Problem 1: cycle with $\mathtt{K}$

$$\Pi_{18} = \{p \leftarrow \mathtt{K}\, p\}$$

- ☹ has two world views $\{\emptyset\}$ and $\{\{p\}\}$ [Gelfond 1991,1994], [Wang&Zhang 2005]
- ☺ has unique world view $\{\emptyset\}$ [Gelfond 2011, Kahl 2014, FHS]

Remark. clear case: $\mathtt{K}\, p \rightarrow p$ is the truth axiom of epistemic logic

Problem 2: cycle with $\mathtt{M}$

$$\Pi_1 = \{p \leftarrow \mathtt{M}\, p\}$$

? has unique world view $\big\{\{p\}\big\}$ [Kahl 2014]

? has 2 world views $\big\{\emptyset\big\}$ and $\big\{\{p\}\big\}$ [Gelfond 1991,1994], [Wang&Zhang 2005]

? has unique world view $\big\{\emptyset\big\}$ [FHS]

has 2 world views $\big\{\emptyset\big\}$ and $\big\{\{p\}\big\}$ if $\mathtt{M}$ replaced by $\neg\mathtt{K}\,\neg$ [FHS]

Remark. circular $\Rightarrow$ no clear intuitions (at least for us)

Problem 3: preference over a disjunction

$$\Pi_{32} = \{p \, \texttt{or} \, q \leftarrow, \; q \leftarrow \texttt{M} \, p\}$$

- ☹ has no world view [Gelfond 1991,1994,2011]
- ☺ has unique world view $\big\{\{q\}\big\}$ [Kahl 2014, FHS]

Remark. intuitively clear (similar to Gelfond's eligibility example)

Problem 4: preference over a disjunction, ctd.

$$\Pi_{32} = \{p \,\texttt{or}\, q \leftarrow, \; q \leftarrow \texttt{not}\, \mathrm{K}\, p\}$$

- ☺ has unique world view $\big\{\{q\}\big\}$ [Kahl 2014]
- ☹ has 2 world views $\big\{\{q\}\big\}$ and $\big\{\{p\}\big\}$ [Gelfond 1991,1994,2011, FHS]

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 = $(\mathcal{W}, H, T)$ where
  - $\mathcal{W}$ is a classical S5 model: $\mathcal{W} \subseteq 2^{\text{PVAR}}$
  - (H, T) is an ht-model: $H \subseteq T \subseteq$ PVAR
    $\Rightarrow$ H and T not necessarily in $\mathcal{W}$ (!)
- truth conditions:

$\mathcal{W}, H, T \models \mathtt{K}\varphi$ iff $\mathcal{W}, H', T' \models \varphi$ for every ht-model H', T' that can be built from $\mathcal{W}$

$\mathcal{W}, H, T \models \mathtt{M}\varphi$ iff $\mathcal{W}, H', T' \models \varphi$ for some ht-model ...

- $\langle \mathcal{W}, T, T \rangle$ is an epistemic equilibrium model of φ iff $\langle \mathcal{W}, T, T \rangle \models \varphi$ and $\langle \mathcal{W}, H, T \rangle \not\models \varphi$ for every $H \subset T$
- $\langle \mathcal{W} \rangle$ is an equilibrium view of φ iff $\mathcal{W}$ is the maximal collection satisfying $\mathcal{W} = \{T : \langle \mathcal{W}, T, T \rangle$ is an epi.eq.model of $\varphi\}$

Theorem (Wang&Zhang 2005, Thm. 2)

$\mathcal{W}$ is a world view of Π iff $\mathcal{W}$ is an equilibrium view of Π.

[Wang & Zhang 2005]'s epistemic extension of HT: criticisms

1. not really an epistemic logic
   - $p \wedge \mathtt{K}\,\neg p$ has a model (and even a WZ-equilibrium model)
2. not really an intuitionistic modal logic
   - $\mathtt{K}\,\varphi \leftrightarrow \neg\mathtt{M}\,\neg\varphi$ valid
   - $\mathtt{K}\,\neg\neg\varphi \rightarrow \mathtt{K}\,\varphi$ valid
   - $\neg\neg\mathtt{K}\,\varphi \rightarrow \mathtt{K}\,\varphi$ valid
3. equilibrium definition unintuitive beyond disjunctive logic programs ('nested epistemic logic programs', NELP)
   - $(\mathcal{W}, T, T)$ is WZ-equilibrium model of $\mathtt{K}\,p$ iff $\mathcal{W}$ S5-model of $\mathtt{K}\,p$ and $T{=}\emptyset$
     $\Rightarrow$ no minimisation
   - $\mathtt{K}\,p$ has no WZ-equilibrium model
   - $\mathtt{M}\,p \wedge \mathtt{M}\,\neg p$ 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
   - $\emptyset \mathrel{|\!\approx}_{EE} \mathtt{K}\,\neg p$
   - $p \vee q \mathrel{|\!\approx}_{EE} \mathtt{K}\,(p \vee q)$
   - $p \vee q \mathrel{|\!\not\approx}_{EE} \mathtt{M}\,p \wedge \mathtt{M}\,q$
3. maximise ignorance: cf. Levesque's "all-that-I-know" and Moore's autoepistemic logic
   - $p \vee q \mathrel{|\!\approx}_{AEE} \mathtt{M}\,p \wedge \mathtt{M}\,q$
   - however makes no difference for the discriminating examples

Our epistemic ht-models

two-dimensional modal logic (cf. intuitionistic S5)

Definition

e-ht-model = $(\mathcal{W}, \hbar)$ where

- $\mathcal{W}$ is a classical S5 model: $\mathcal{W} \subseteq 2^{\mathrm{PVAR}}$
- $\hbar : \mathcal{W} \longrightarrow 2^{\mathrm{PVAR}}$ such that $\hbar(T) \subseteq T$ for every $T \in W$

- classical S5 model: $\hbar = id$
- truth conditions:

$$
\begin{array}{lll}
(\mathcal{W},\hbar), T \models p & \text{iff} & p \in \hbar(T) \\
(\mathcal{W},\hbar), T \models \varphi \to \psi & \text{iff} & (\mathcal{W},\hbar), T \models \varphi \supset \psi \text{ and} \\
 & & (\mathcal{W}, id), T \models \varphi \supset \psi \\
(\mathcal{W},\hbar), T \models \mathtt{K}\,\varphi & \text{iff} & (\mathcal{W},\hbar), T' \models \varphi \text{ for every } T' \in \mathcal{W} \\
(\mathcal{W},\hbar), T \models \mathtt{M}\,\varphi & \text{iff} & (\mathcal{W},\hbar), T' \models \varphi \text{ for some } T' \in \mathcal{W}
\end{array}
$$

- satisfies the requirements for intuitionistic modal logics [Fischer-Servi 1976, Fariñas&Raggio 1983, Simpson 1995, ...]

Our epistemic equilibrium models

minimise truth (cf. equilibrium logic)

Definition

$\mathcal{W}$ is an epistemic equilibrium model of φ iff

1. $(\mathcal{W}, id), T \models \varphi$ for every $T \in \mathcal{W}$ (classical S5 model of φ)
2. there is no $\hbar \neq id$ such that $(\mathcal{W}, \hbar), T \models \varphi$ for every $T \in \mathcal{W}$ (no 'weaker' e-ht-model of φ)

Example: $\{p \;\mathtt{or}\; \overline{p} \leftarrow\}$ has 3 epistemic eq.models:

$$\{\emptyset\},\ \{\{p\}\},\ \text{and}\ \{\emptyset, \{p\}\}$$

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 φ iff

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

Example: $\{p \;\texttt{or}\; \overline{p} \leftarrow\}$ has 1 autoepistemic eq.model:

$$\big\{\emptyset, \{p\}\big\}$$

Theorem (strong equivalence)

...

Ongoing work: first minimise knowledge, then truth?

- given Π,
  1. compute the biggest S5 model $\mathcal{W}$ of Π
  2. compute the biggest subset of $\mathcal{W}$ that is an epistemic eq.model
- gets right all the examples but $p \leftarrow \mathtt{M}\, p$

To sum it up

- many possible semantics of epistemic specifications
- arguably flawed: [Gelfond 1991,1994; Wang&Zhang 2005]
- 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