Section 4

Autoepistemic Logic

Subsection 4.1

Introducing autoepistemic logic

Introducing autoepistemic logic: an example

autoepistemic: reflection upon self-knowledge

Idea: formalism to model how an agent forms her own belief sets and how to reason about it.

Example:
Are the Stones playing in Newcastle next week?
No, because otherwise I would have heard about it.

Observations:

- no definite knowledge that the Stones do not give a concert in Newcastle next week.
- incomplete knowledge and negative answer is rather a conjecture

New knowledge: the Stones are giving a concert in Newcastle next week!

Observations:

- old conclusion by introspection is no longer valid and must be revised—nonmonotonic reasoning!
- long-term knowledge ("If something important is to happen in my city, then I would know about it") has not changed.
- what has changed is that answer is based on fact, not on introspection

Introducing autoepistemic logic: another example

Indicate "believed knowledge" by a modal operator L applied to FOL sentences.

$L\varphi$ means intuitively: "I know φ".

Capture:

- Prof Jones is a university professor and thus normally teaches.
- If I do not believe that Dr. Jones does not teach, then Dr. Jones does teach

by the modal formula:

$$Lprof_J \land \neg L \neg teaches_J \longrightarrow teaches_J$$

The concert example can be captured by:

- $concert \longrightarrow Lconcert$ ("If a concert takes place, then I know about it. ")
- $\neg Lconcert$ ("I don't know that a concert takes place. ")

Towards syntax and semantics

The L-operator can appear nested in formulas: $LL\varphi$ or $L\neg Lq$ or $\neg L(p \vee Lq)$

The meaning of autoepistemic logic is given in terms of expansions, i.e., pieces of knowledge defining "world views" compatible with and based on the given knowledge.

Expansions are stable, if

- if fact φ is in an expansion, then so is $L\varphi$
- if fact φ is not in an expansion, then $\neg L\varphi$ is in the expansion

Syntax of autoepistemic logic

Definition 4.1 (Autoepistemic formulas, AE-formula)

Autoepistemic formulas (AE-formulas) are the smallest set satisfying the following:

- each closed FOL formula is an AE-formula
- if φ is an AE-formula, then $L\varphi$ is an AE-formula
- if φ and ψ are AE-formulas, then so are the following:
  - $\neg\varphi$
  - $(\varphi \wedge \psi)$
  - $(\varphi \vee \psi)$
  - $(\varphi \longrightarrow \psi)$

The set of all AE-formulas is denoted by *For*.
An autoepistemic theory (AE-theory) is a set of AE-formulas.

Syntax of autoepistemic logic—schema

Sometimes it is convenient to use open FOL formulas in the scope of the L-operator. In such cases the AE-formula reads as a schema, i.e., a collection of ground instances.

E.g.:

$$\begin{gathered} german(X) \wedge \neg L \neg drinksBeer(X) \longrightarrow drinksBeer(X), \\ german(bob),\ german(lisa) \end{gathered}$$

is read as the autoepistemic theory:

$$\begin{gathered} german(bob) \wedge \neg L \neg drinksBeer(bob) \longrightarrow drinksBeer(bob) \\ german(lisa) \wedge \neg L \neg drinksBeer(lisa) \longrightarrow drinksBeer(lisa) \\ german(bob),\ german(lisa) \end{gathered}$$

Sub-formula

Let φ be an AE-formula. The set of subformulas of φ ($sub(\varphi)$) is defined as:

- $sub(\varphi) = \emptyset$ for FOL formula φ
- $sub(\neg\varphi) = sub(\varphi)$
- $sub(\varphi \vee \psi) = sub(\varphi \wedge \psi) = sub(\varphi \longrightarrow \psi) = sub(\varphi) \cup sub(\psi)$
- $sub(L\varphi) = \{\varphi\}$

Let T be an AE-theory. The set of subformulas of T is defined as

$$sub(T) = \bigcup_{\varphi \in T} sub(\varphi).$$

Note: we do not go further into the structure of a formula, after the out-most occurrence of L.

For example: If $T = \{L\neg Lq, L(Lp \wedge r), \neg Lr, s\}$, then $sub(T) = \{\neg Lq, (Lp \wedge r), r\}$

Some auxiliary notions—degree, kernel

degree

The degree of an AE-formula φ ($degree(\varphi)$) is the maximal depth of L-nestings that occurs in φ.

Let T be an AE-theory, then T_n denotes the set of AE-formulas in T with degree less or equal n.

For example: $degree((\neg L \neg L(p \wedge Lq))) = 3.$

kernel

The kernel of an AE-theory T is defined as the set of all FOL formulas that are elements of T (denoted T_0).

For example: if $T = \{p, \neg Lq, \neg Lq \longrightarrow s, L\neg Lr, r\}$, then $T_0 = \{p, r\}$.

Normal form for autoepistemic formulas

Definition 4.2 (Normal form)

An AE-formula is in normal form, if it has the form

$$\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_n,$$

where each φ_i has the form

$$\beta \vee L\gamma_1 \vee \cdots \vee L\gamma_p \vee \neg L\delta_1 \vee \cdots \vee \neg L\delta_q$$

with a FOL formula β.

Each AE-formula φ can be transformed into an equivalent AE-formula ($nf(\varphi)$) in normal form, such that $degree(\varphi) = degree(nf(\varphi))$.

Definition 4.3 (AE-interpretation)

An autoepistemic interpretation $\mathcal{I}$ over a signature Σ provides

- a non-empty domain $\mathit{dom}(\mathcal{I})$
- an interpretation $f^{\mathcal{I}}$ for each function symbol $f \in \Sigma$ (as in FOL)
- an interpretation $r^{\mathcal{I}}$ for each predicate symbol $r \in \Sigma$ (as in FOL)
- a truth value $(L\varphi)^{\mathcal{I}}$ for every AE-formula $L\varphi$.

As in FOL, $\mathcal{I} \models \varphi$ indicates that an AE-interpretation $\mathcal{I}$ satisfies an AE-formula (is an AE-model of) φ.

A formula logically follows from a set M of AE-formulas ($M \models \varphi$) iff φ is valid in all AE-models of M.

For a set of AE-formulas M, $\mathit{Th}(M)$ is the set of AE-formulas that logically follow from M.

Remarks on the semantics

In Def. 4.3 the validity of φ in $\mathcal{I}$ and the validity of $L\varphi$ in $\mathcal{I}$ are unrelated:
$L\varphi$ is treated as a new atom (a 0-ary predicate) and thus independent of φ.

Intuition:
φ expresses truth of φ, whereas $L\varphi$ expresses belief in (/knowledge of) φ.

This choice of semantics admits to
"believe in something false", or "not to believe in something true".

The following alternative definition of the semantics captures this observation.

Algebra-based semantics of autoepistemic logics

An algebra with a belief set is a pair $(\mathcal{B}, \mathit{Bel})$, where

- $\mathcal{B}$ is a first order interpretation and
- Bel is a set of AE-formulas.

Validity of AE-formulas in $(\mathcal{B}, \mathit{Bel})$ is defined as:

- $(\mathcal{B}, \mathit{Bel}) \models \varphi$ iff $\mathcal{B} \models \varphi$ for a closed FO formula φ
- $(\mathcal{B}, \mathit{Bel}) \models \neg\varphi$ iff $(\mathcal{B}, \mathit{Bel}) \not\models \varphi$
- $(\mathcal{B}, \mathit{Bel}) \models (\varphi \vee \psi)$ iff $(\mathcal{B}, \mathit{Bel}) \models \varphi$ or $(\mathcal{B}, \mathit{Bel}) \models \psi$
- $(\mathcal{B}, \mathit{Bel}) \models (\varphi \wedge \psi)$ iff $(\mathcal{B}, \mathit{Bel}) \models \varphi$ and $(\mathcal{B}, \mathit{Bel}) \models \psi$
- $(\mathcal{B}, \mathit{Bel}) \models (\varphi \longrightarrow \psi)$ iff $(\mathcal{B}, \mathit{Bel}) \models \varphi$ implies $(\mathcal{B}, \mathit{Bel}) \models \psi$
- $(\mathcal{B}, \mathit{Bel}) \models L\varphi$ iff $\varphi \in \mathit{Bel}$.

Relationship between the two semantics

The semantics are equivalent.

1. From a given AE-interpretation $\mathcal{I}$, we define an algebra with a belief set $(\mathcal{B}, Bel)$ as follows:
   - the domain of $\mathcal{B}$ and the interpretation function of predicate and function symbols are same as in $\mathcal{I}$.
   - $Bel = \{\varphi \mid (L\varphi)^{\mathcal{I}} = true\}$
2. From a given algebra with a belief set $(\mathcal{B}, Bel)$, we define an AE-interpretation $\mathcal{I}$ as follows:
   - the domain of $\mathcal{I}$ and the interpretation function of predicate and function symbols are same as in $\mathcal{B}$.
   - $(L\varphi)^{\mathcal{I}} = true$ iff $\varphi \in Bel$.

Convention: We use the two semantics interchangeably.
By "an AE-interpretation with belief set *Bel*" we mean $Bel = \{\varphi \mid (L\varphi)^{\mathcal{I}} = true\}$.

We define "φ follows from AE-theory T w.r.t. belief set E" (denoted $T \models_E \varphi$) as φ is valid in every AE-model of T with belief set E.

Subsection 4.2

Expansions of autoepistemic theories

Towards expansions — considerations

What knowledge would an agent with introspection have
given a set of facts (i.e. AE-formulas) T?

The agent's knowledge would be a set E of AE-formulas that

- includes T
- allows introspection
- is grounded in T
  (meaning: the knowledge in E must be reconstructable from:
  T, belief in (knowledge of) E, and non-belief in (non-knowledge of) E)

Let T and E be sets of AE-formulas. We define the following sets

- $LE = \{L\varphi \mid \varphi \in E\}$
- $\neg LE^C = \{\neg L\psi \mid \psi \notin E\}$
- $\Omega_T(E) = \{\varphi \mid T \cup LE \cup \neg LE^C \models \varphi\}$

Definition 4.4 (Expansion)

Let T and E be sets of AE-formulas.

- E is T-sound iff $E \subseteq \Omega_T(E)$
- E is T-complete iff $\Omega_T(E) \subseteq E$
- E is an expansion of T iff $E = \Omega_T(E)$

Intuition:
The agent decides to believe in a set of AE-formulas T.
Based on this, a set of AE-formulas can be deduced from T and the beliefs adopted ($LE \cup \neg LE^C$). If the deduced set is exactly the set of beliefs E, then E is an expansion.

Alternative characterization of expansions

Observation:
AE-models of $T \cup LE \cup \neg LE^C$ are just the AE-models of T with belief set E!

Thus we obtain an alternative characterization of expansions.

Corollary 4.5

E is an expansion of an AE-theory T iff $E = \{\varphi \mid T \models_E \varphi\}$.

Example 4.6

Consider the AE-theory T_1:

$$\{german \wedge \neg L \neg drinksBeer) \longrightarrow drinksBeer, \; german\}$$

This AE-theory has one expansion.

The formula $\neg drinksBeer$ cannot be derived before $\neg L \neg drinksBeer$ is contained in the expansion.

The only expansion of T_1 has the kernel: $Th(\{german, drinksBeer\})$

If we extend T_1 by adding:

$$\{(eatsPizza \wedge \neg L drinksBeer) \longrightarrow \neg drinksBeer, \; eatsPizza\}$$

then the theory has two expansions:

- kernel of the first expansion: $\{geman, eatsPizza, drinksBeer\}$
- kernel of the second expansion: $\{geman, eatsPizza, \neg drinksBeer\}$

Subsection 4.3

Stable sets and their properties

Stable sets — origin

- Stable belief sets were introduced by Robert Stalnaker in the early '80s
- proposed as a formal representation of the epistemic state of an ideally rational agent, with full introspective capabilities.
- Assumes a propositional language, endowed with a modal operator $\Box\varphi$ interpreted as "φ is believed"
- a set of formulas is a stable set if it is "stable" under classical inference and epistemic introspection
- influenced research on AE logics and nonmonotonic logics in general

Definition 4.7 (stable sets)

Let E be a set of autoepistemic formulas. E is called stable iff

- E is deductively closed, i.e. $E = Th(E)$,
- $\varphi \in E$ implies $L\varphi \in E$, for all AE-formula φ, and
- $\varphi \notin E$ implies $\neg L\varphi \in E$, for all AE-formula φ

Note: Expansions are stable sets by definition.
Thus they inherit all the properties we show for stable sets.

Theorem 4.8

For an AE-theory T and a set of AE-formulas E the following statements are equivalent:

1. *E is an expansion of T*
2. *E is stable, $T \subseteq E$ and is T-sound.*

Proof: blackboard

Lemma 4.9

For a stable set E and an AE-formula φ the following statements are equivalent:

a) $E \models_E \varphi$

b) $E \models \varphi$

c) $\varphi \in E$

For a FOL formula φ, the statements a)-c) are equivalent to

d) $E_0 \models \varphi$

Proof: blackboard

Stable sets are determined by their kernels

Stable sets are uniquely determined by their objective subsets, i.e. their kernels.

Theorem 4.10

For stable sets E and F, $E_0 = F_0$ implies $E = F$.

Proof: blackboard

Existence of stable sets

How can expansions be computed? A first hint

Theorem 4.11

Let T be a first order theory. Then there is a stable set E with $E_0 = T$.

Proof: blackboard

Theorem 4.12 (Orthogonality of stable sets)

Let E and F be different stable sets. Then $E \cup F$ is inconsistent.

Proof: blackboard

Theorem 4.13

If E is a stable set then it is an expansion of E_0.

Proof: blackboard

Subsection 4.4

Computing expansions of AE-theories

Considerations

To achieve nonmonotonic behavior w.r.t. AE-theories, formulas ("conjectures") can be added to the set of believes that need not be added.

What makes computing expansions difficult?

- nested occurrences of the L-operator
- infinitely many conjectures. How to compute all expansions?

How to remedy this?

- Nested occurrences of L-operator: concentrate on potential kernels of expansions (Theorem 4.10).
- by Coincidence Lemma: it suffices to consider beliefs or non-beliefs in formulas from $sub(T)$ to determine the expansions of T.

  Only those formulas with L-operator play a role in the interpretation of T.

Overview of the computation procedure for expansions

Compute expansions of AE-theories by:

- partition $sub(T)$ into:
  - $E(+)$: set of beliefs
  - $E(-)$: set of non-beliefs
- Compute the corresponding kernel $E(0)$ of a potential expansion, using T, beliefs in $E(+)$ and non-beliefs in $E(-)$.
- Check whether the stable set determined by $E(0)$ is indeed an expansion

Example 4.14

Let $T = \{Lp \longrightarrow p\}$.
Since $Lp \longrightarrow p$ is the only AE-formula occurring (at top-level) of T, $sub(T) = \{p\}$.
There are two partitions of $sub(T) = \{p\}$.

$E(+)$	$E(-)$	$E(0)$	$E(+) \subseteq E(0)$?	$E(-) \cap E(0) = \emptyset$?	expansion?
$\{p\}$	$\emptyset$	$Th(\{p\})$	yes	yes	yes
$\emptyset$	$\{p\}$	$Th(\emptyset)$	yes	yes	yes

- $E(0)$: set of first order formula that follow from T.
- condition $E(+) \subseteq E(0)$:
  test whether everything that the agent believes in is in $E(0)$.
- condition $E(-) \cap E(0) = \emptyset$:
  ensures that $E(0)$ does not include non-beliefs of the agent

Procedure for computing expansions for AE-theories without L-nestings

Compute expansions no L-nesting (T)

$Expansions := \emptyset$
for all partitions $E(+)$ and $E(-)$ of $sub(T)$ **do**
    $E(0) := \{\varphi \in For_0 \mid T \cup LE(+) \cup \neg LE(-) \models \varphi\}$
    if $E(+) \subseteq E$ AND $E(-) \cap E = \emptyset$ **then**
        $Expansions := Expansions \cup \{E(0)\}$
    end if
end for
return $Expansions$

Example 4.15

Let $T = \{Lp \longrightarrow p, \neg L\neg Lp\}$, with $sub(T) = \{p, \neg Lp\}$.
Now the partitions of $sub(T)$ are no longer first order formulas!

$E(+)$	$E(-)$	$E(0)$	$E(+) \subseteq E$?	$E(-) \cap E = \emptyset$?	expansion?
$\{p, \neg Lp\}$	$\emptyset$	For_0	yes	yes	yes
$\{p\}$	$\{\neg Lp\}$	$Th(\{p\})$	yes	yes	yes
$\{\neg Lp\}$	$\{p\}$	For_0	yes	no	no
$\emptyset$	$\{p, \neg Lp\}$	$Th(\emptyset)$	yes	no	no

- Line 1: $E(0)$ is inconsistent, since $L\neg Lp$ follows from $LE(+)$, but $\neg L\neg Lp \in T$.
- Line 2: $T \cup LE(+) \cup \neg LE(-) = \{Lp \longrightarrow p, \neg L\neg Lp, Lp\}$, thus $E(0) = Th(\{p\})$. Since $p \in E$ and E is stable and consistent, we have $Lp \in E$ and thus $\neg L \notin E$.
- Line 3: $T \cup LE(+) \cup \neg LE(-)$ contains both $L\neg Lp$ and $\neg L\neg Lp$, thus $E(0) = For_0$
- Line 4: $T \cup LE(+) \cup \neg LE(-) = \{Lp \longrightarrow p, \neg L\neg Lp, \neg Lp\}$. From $p \notin E$ follows $\neg Lp \in E$ and thus $E(-) \cap E \neq \emptyset$

Procedure for computing expansions for general AE-theories

Compute expansions (T)

$\mathit{Expansions} := \emptyset$
for all partitions $E(+)$ and $E(-)$ of $\mathit{sub}(T)$ **do**
    $E(0) := \{\varphi \in \mathit{For}_0 \mid T \cup LE(+) \cup \neg LE(-) \models \varphi\}$
    Let E be the unique stable set with kernel $E(0)$
    if $E(+) \subseteq E$ AND $E(-) \cap E = \emptyset$ **then**
        $\mathit{Expansions} := \mathit{Expansions} \cup \{E(0)\}$
    end if
end for
return $\mathit{Expansions}$

Towards the correctness proof

Lemma 4.16 (Preservation Lemma)

Let E be a stable set and T an AE-theory.
If $E_0 = \{\varphi \in For_0 \mid T \cup LE \cup \neg LE^C \models \varphi\}$, then $E = \{\varphi \in For \mid T \cup LE \cup \neg LE^C \models \varphi\}$.

Proof: blackboard

Lemma 4.17 (Coincidence Lemma)

Let T be an AE-theory. Consider sets of AE-formulas $E(+)$, $E(-)$, $F(+)$, and $F(-)$ with the following properties:

- $sub(T) \subseteq E(+) \cup E(-)$ *and* $E(+) \cap E(-) = \emptyset$ *and* $sub(T) \subseteq F(+) \cup F(-)$ *and* $F(+) \cap F(-) = \emptyset$
- $E(+) \cap sub(T) = F(+) \cap sub(T)$
- $E(-) \cap sub(T) = F(-) \cap sub(T)$.

Then the same first order formula follow from $T \cup LE(+) \cup \neg LE(-)$ as from $T \cup LF(+) \cup \neg LF(-)$

Proof: blackboard

Theorem 4.18

Let T be an AE-theory and let $sub(T)$ be partitioned into the disjoint sets $E(+)$ and $E(-)$. We consider the following steps:

1. *Compute $E_0 = \{\varphi \in For_0 \mid T \cup LE(+) \cup \neg LE(-) \models \varphi\}$ and let E be the unique stable set with kernel E_0.*
2. *Check whether $E(+) \subseteq E$ and $E(-) \cap E = \emptyset$.*

Then the following holds:

a) *If the check in Step 2. is positive, then E is an expansion of T.*

b) *Conversely, for every expansion E of T there is a decomposition of $sub(T)$ into $E(+)$ and $E(-)$ such that*
   - *$E(0) = E_0$ and*
   - *the check in Step 2 is positive.*

Proof: blackboard

Subsection 4.5

Embedding Default Logic into AE-Logic

Default logic vs. autoepistemic logic

How to embed default logic into autoepistemic logic?

- Default logic: uses rules
  AE-logic: uses introspection

- With $L\varphi$ means "φ is known", we get:

$$\frac{\mathit{german} : \mathit{drinksBeer}}{\mathit{drinksBeer}} \qquad \text{vs.} \qquad L\mathit{german} \land \neg L\neg\mathit{drinksBeer} \longrightarrow \mathit{drinksBeer}$$

What is the semantic relationship between the two formalisms?

Translating default theories to AE-theories

Idea: express consistency of justifications ψ by $\neg L \neg \psi$ ("$\neg\psi$ is not known")

Definition 4.19 (*trans*())

Let $\delta = \frac{\varphi : \psi_1, \ldots, \psi_n}{\chi}$ be a default rule. We define the translation function for default rules as follows:

$$trans(\delta) = L\varphi \wedge \neg L \neg \psi_1 \wedge \cdots \wedge \neg L \neg \psi_n \longrightarrow \chi.$$

Let $T = (W, D)$ be a default theory. We define the translation function for default theories as follows:

$$trans(T) = W \cup \{trans(\delta) \mid \delta \in D\}.$$

Does this translation preserve the semantics?

How to compare the semantics?

Recall:

- Extension of a default theory: FO formulas only
- Expansion of a AE-theory: FO formulas possibly in scope of L-operator

Approach for comparison:

Compare extensions of default theory T with kernels of expansions of translated formulas $trans(T)$.
—such kernels are unique (see Section 4.3) and FO formulas

Plan for this section:
In the following we want to derive conditions under which extensions of a default theory and expansions (of the translated default theory coincide).

Example – difference of expansions and extensions

Example 4.20

Consider the default theory $T_{ex1} = (W, D)$ with $W = \emptyset$ and $D = \left\{ \frac{p : true}{p} \right\}$
The translation is $trans(T_{ex1}) = \{Lp \wedge \neg Lfalse \longrightarrow p\}$

The only extension of T_{ex1} is $Th(\emptyset)$,
but $trans(T_{ex1})$ has two expansions: $Th(\emptyset)$ and $Th(\{p\})$.

The second expansion comes from the self-referential nature of expansions!

$$E = \{\varphi \mid T \cup LE \cup \neg LE^C \models \varphi\}$$

If it is decided to believe in p (and not in *false*), then p can be derived!
Whereas in default logic p needs to be known by other information!

Definition 4.21

Let T be an AE-theory and E an expansion of T. E is an AE-minimal expansion of T iff there is no expansion F of T s.t. $F_0 \subset E_0$.

The idea is to concentrate on those expansions (that include the theory and) that cannot be "generated" from a smaller kernel in size.

Does it help?

The AE-theory $\mathit{trans}(T_{ex1})$ from Example 4.20 has one AE-minimal expansion with the kernel: $\mathit{Th}(\emptyset)$ which is the extension of T_{ex1}.

Example 4.22

Consider the default theory $T_{ex2} = (W, D)$ with $W = \emptyset$ and $D = \left\{ \frac{true : \neg p}{q}, \frac{p : true}{p} \right\}$ which has the single extension $Th(\{q\})$.

The AE-theory $trans(T_{ex2}) = \{Ltrue \wedge \neg L\neg\neg p \longrightarrow q, Lp \wedge \neg Lfalse \longrightarrow p\}$ has two expansions:

- $\widehat{E}$ with kernel $\widehat{E}_0 = Th(\{q\})$ and
- $\widehat{F}$ with kernel $\widehat{F}_0 = Th(\{p\})$

Both expansions are AE-minimal. But the set of expansions does not coincide with the extension of T_{ex2}.

AE-minimality still admits to deliberately believe in Lp.

Addressing groundedness of expansions:
avoiding arbitrary formulas in expansions by restricting self-referentiality.

Definition 4.23 (SS-minimal)

Let T be an AE-theory and E an expansion of T.
E is an SS-minimal expansion of T iff there is no stable set F s.t. $T \subseteq F$ and $F_0 \subset E_0$.

SS-minimality implies AE-minimality, but the converse does not hold.

Restricting AE-interpretations to those with stable belief sets:

Definition 4.24

Let SS denote the class of all stable sets.
We define $T \models_{SS} \varphi$ iff $T \models_E \varphi$ for all stable sets E.

FO self-referentiality of expansions

Since $\models_{SS}$ is stronger than $\models$, it allows us to weaken the premises used in the definition of an expansion without losing information.

Lemma 4.25

A set of AE-formulas E is an expansion of an AE-theory T iff
$E = \{\varphi \mid T \cup LE_0 \cup \neg L(For_0 \setminus E_0) \models_{SS} \varphi\}$.

Proof: exercise

Intuition of Lemma 4.25 is that the self-referentiality in the definition of expansions has been restricted to FO beliefs.

Moderately grounded expansions

Observation:
Since the only beliefs admitted are those in T, it is admissible to replace E_0 by T in Lemma 4.25.

Definition 4.26

E is a moderately grounded expansion of an AE-theory T iff
$E = \{\varphi \mid T \cup LT \cup \neg L(For_0 \setminus E_0) \models_{SS} \varphi\}$.

Lemma 4.27

Let T be an AE-theory and E a set of AE-formulas.
E is a moderately grounded expansion iff E is a SS-minimal expansion of T.

Restricting expansions: grounding expansions

Do SS-minimal expansions and extensions coincide?

Consider the default theory T_{ex2} from Example 4.22 again.
Recall: the expansions $\widehat{E}$ with kernel $Th(\{q\})$ and $\widehat{F}$ with kernel $Th(\{p\})$ are AE-minimal.

They are also SS-minimal:

- Let S be a stable set with $S \subseteq T$. Suppose $S_0 \subset \widehat{F}_0$,
  then $p \notin S$, $\neg\neg p \notin S$, thus $\neg L \neg\neg p \in S$ and so $q \in S$ and $q \in S_0$.
  But then $S_0 \not\subset \widehat{F}_0$, which is a contradiction.
- Suppose $S_0 \subset \widehat{E}_0$, then $p \notin S$ and $\neg\neg p \notin S$,
  therefore $\neg L \neg\neg p \in S$, and so $q \in S$ and $q \in S_0$.
  Since S is deductively closed, S_0 is deductively closed, too.
  Since $q \in S$, S_0 is nor a proper subset of $\widehat{E}_0$, which is a contradiction.

Analyzing Example 4.22

Recall: $\mathit{trans}(T_{ex2}) = \{L\mathit{true} \wedge \neg L\neg\neg p \longrightarrow q, Lp \wedge \neg L\mathit{false} \longrightarrow p\}$ and
its expansion $\widehat{F}$ has kernel $\widehat{F}_0 = Th(\{p\})$.

How was p derived from $\big(\mathit{trans}(T_{ex2}) \cup L(\mathit{trans}(T_{ex2})) \cup \neg L(For_0 \setminus \widehat{F}_0)\big)$?

Let $(\mathcal{I}, S)$ be an AE-model of $(\mathit{trans}(T_{ex2}) \cup L(\mathit{trans}(T_{ex2})) \cup \neg L(For_0 \setminus \widehat{F}_0))$ with stable set S. Then $L(\neg L\neg\neg p \longrightarrow q) \in S$.

By stability and consistency of S: $\neg L\neg\neg p \longrightarrow q \in S$. So, $L\neg\neg p \in S$ or $\{q, Lq\} \subseteq S$.

Since $q \notin \widehat{F}_0$, $(\mathcal{I}, S) \models \neg Lq$ and thus $q \notin S$ holds.

We can conclude: $L\neg\neg p \in S$, thus $\neg\neg p \in S, p \in S$ and $(\mathcal{I}, S) \models Lp$.

Using $(\mathcal{I}, S) \models Lp \wedge \neg L\mathit{false} \longrightarrow p$ and $(\mathcal{I}, S) \models \neg L\mathit{false}$, we finally get $(\mathcal{I}, S) \models p$.

Note that $L\neg\neg p$ was obtained before $\neg\neg p$ (self-referential still!).
Formula $L\neg\neg p$ was obtained from rule $\neg L\neg\neg p \longrightarrow Lq$. It was applied using contraposition, i.e. $\neg Lq \longrightarrow L\neg\neg p$.
But, the corresponding default $\frac{\mathit{true} : \neg p}{q}$ can only be used from top to bottom!

Restricting expansions: enforcing unidirectional application

Addressing the possibility to apply AE-implications in both directions.

AE-formula in default normal form are AE-formulas
$L\varphi \wedge \neg L\neg\psi_1 \wedge \cdots \wedge \neg L\neg\psi_n \longrightarrow \chi$, where $\varphi, \psi_1, \ldots \psi_n, \chi$ are FO formulas.[6]

Definition 4.28

Let T be a AE-theory consisting of FO formulas and AE-formulas in default normal form and let E be an expansion of T.

T^E denotes the set of AE-formulas $L\varphi \wedge \neg L\neg\psi_1 \wedge \cdots \wedge \neg L\neg\psi_n \longrightarrow \chi$ in T such that $\neg\psi_i \notin E$ (for $1 \leq i \leq n$).

E is strongly grounded in T iff the following holds:

$$E = \{\varphi \mid T^E \cup LT^E \cup \neg L(For_0 \setminus E_0) \models_{SS} \varphi\}.$$

For a strongly grounded expansion E it is impossible to obtain $L\psi_i$ from not knowing the consequent χ!

---

[6] They are the translations of defaults into AE-logic.

Strongly grounded implies moderately grounded

Lemma 4.29

Let E be a strongly grounded expansion of an AE-theory T.
Then E is a moderately grounded (and thus SS-minimal) expansion of T.

Proof: blackboard.

Difference between expansions and extensions

To sum up:
Expansions vs. extensions

1. Expansions are not necessarily minimal w.r.t. kernel inclusion. Extensions cannot be subsets of other extensions (of the same default theory).
2. Expansions may not be "well-grounded" in the given knowledge; can include AE-formulas that it was decided to believe in.
3. AE-formulas may be used in both directions, whereas default rules are strictly unidirectional.

Extensions and strongly grounded expansions coincide

Theorem 4.30

Let $T = (W, D)$ be a default theory.
For every extension E of T there is a strongly grounded expansion F of $\mathit{trans}(T)$ such that $E = F_0$.
Conversely, the kernel of every strongly grounded expansion of $\mathit{trans}(T)$ is an extension of T.

Computational complexity of reasoning in autoepistemic logics

- For closed FOL formulas in a logic L holds: if satisfiability in L is decidable, then so are nonmonotonic reasoning tasks for L.
- deciding whether an AE-theory has a stable expansion: Σ_2^P-complete
- credulous reasoning is Σ_2^P-complete
  cautious reasoning is Π_2^P-complete