Human Reasoning, Logic Programs and the CORE Method

Steffen Hölldobler
International Center for Computational Logic
Technische Universität Dresden
Germany

- Human Reasoning
- Three-Valued Logic Programs
- A CORE Method for Human Reasoning
- Contractions
- Open Problems

“Logic is everywhere...”
Human Reasoning – Two Examples

 ► Instructions on boarding card distributed at Amsterdam Schiphol Airport:
   ▶ If it’s thirty minutes before your flight departure, make your way to the gate.
   As soon as the gate number is confirmed, make your way to the gate.

 ► Notice in London Underground:
   ▶ If there is an emergency then you press the alarm signal bottom.
   The driver will stop if any part of the train is in a station.

 ► Observations
   ▶ Intended meaning differs from literal meaning.
   ▶ Rigid adherence to classical logic is no help in modeling the examples.
   ▶ There seems to be a reasoning process towards more plausible meanings.

   ► The driver will stop the train in a station
   if the driver is alerted to an emergency
   and any part of the train is in the station (see Kowalski 2010).
Human Reasoning – Forward Reasoning

- If she has an essay to write then she will study late in the library.
  She has an essay to write.
  - Modus Ponens.
  - Byrne 1989 96% of subjects conclude that she will study late in the library.

- If she has an essay to write then she will study late in the library.
  She has an essay to write.
  If she has a textbook to read she will study late in the library.
  - Alternative Arguments.
  - Byrne 1989 96% of subjects conclude that she will study late in the library.

- If she has an essay to write then she will study late in the library.
  She has an essay to write.
  If the library stays open she will study late in the library.
  - Byrne 1989 38% of subjects conclude that she will study late in the library.
  - Additional arguments lead to suppression of earlier conclusions.
Reasoning Towards an Appropriate Logical Form

▶ Context independent rules
  ▶ If she has an essay to write and the library is open then she will study late in the library.

▶ Context dependent rule plus exception
  ▶ If she has an essay to write then she will study late in the library.
    However, if the library is not open, then she will not study late in the library.
  ▶ The last sentence is the contrapositive of the converse of the original sentence!
Human Reasoning – Denial of Antecedent

► If she has an essay to write then she will study late in the library.
She does not have an essay to write.
▶ Byrne 1989 46% of subjects conclude that she will not study late in the library.

► If she has an essay to write then she will study late in the library.
She does not have an essay to write.
If she has a textbook to read she will study late in the library.
▶ Byrne 1989 4% of subjects conclude that she will not study late in the library.

► If she has an essay to write then she will study late in the library.
She does not have an essay to write.
If the library stays open she will study late in the library.
▶ Byrne 1989 63% of subjects conclude that she will not study late in the library.
Human Reasoning – The Search for Models

► **Goal** find a logic which adequately models human reasoning.
► How about classical two-valued propositional logic?
► Let’s consider a direct encoding:

\[
\begin{align*}
\{ & l \leftarrow e, e \\
\} & l \leftarrow e, e, l \leftarrow t \\
\} & l \leftarrow e, e, l \leftarrow o \\
\{ & l \leftarrow e, \neg e \\
\} & l \leftarrow e, \neg e, l \leftarrow t \\
\} & l \leftarrow e, \neg e, l \leftarrow o
\end{align*}
\]
Two-Valued Interpretations

- Let $\mathcal{L}$ be a language of propositional logic.
- A (two-valued) interpretation is a mapping $\mathcal{L} \mapsto \{\top, \bot\}$ represented by $I$, where $I$ is a set containing all atoms which are mapped to $\top$.
  - All atoms which do not occur in $I$ are mapped to $\bot$.
- Let $\mathcal{I}$ denote the set of all interpretations.
  - $(\mathcal{I}, \subseteq)$ is a lattice.
- An interpretation $I$ is a model for a program $\mathcal{P}$, in symbols $I \models \mathcal{P}$, iff $I(\mathcal{P}) = \top$.

| $\emptyset$ | $\not\models \{l \leftarrow e, e\}$ |
| $\{e\}$ | $\not\models \{l \leftarrow e, e\}$ |
| $\{l\}$ | $\not\models \{l \leftarrow e, e\}$ |
| $\{e, l\}$ | $\models \{l \leftarrow e, e\}$ |
Two-Valued Interpretations

- Let $\mathcal{L}$ be a language of propositional logic.
- A (two-valued) interpretation is a mapping $\mathcal{L} \mapsto \{ \top, \bot \}$ represented by $I$, where $I$ is a set containing all atoms which are mapped to $\top$.
  - All atoms which do not occur in $I$ are mapped to $\bot$.
- Let $\mathcal{I}$ denote the set of all interpretations.
  - $(\mathcal{I}, \subseteq)$ is a lattice.
- An interpretation $I$ is a model for a program $\mathcal{P}$, in symbols $I \models \mathcal{P}$, iff $I(\mathcal{P}) = \top$.

\[
\begin{align*}
\emptyset & \models \{ l \leftarrow e, \neg e \} \\
\{ l \} & \models \{ l \leftarrow e, \neg e \} \\
\{ e \} & \not\models \{ l \leftarrow e, \neg e \} \\
\{ e, l \} & \not\models \{ l \leftarrow e, \neg e \}
\end{align*}
\]
Logical Consequence (1)

A formula $G$ is a logical consequence of a set of formulas $\mathcal{F}$, in symbols $\mathcal{F} \models G$, iff all models for $\mathcal{F}$ are also models for $G$.

\[
\{l \leftarrow e, e\} \models l \\
\{l \leftarrow e, e\} \models e
\]
Logical Consequence (2)

- A formula $G$ is a logical consequence of a set of formulas $\mathcal{F}$, in symbols $\mathcal{F} \models G$, iff all models for $\mathcal{F}$ are also models for $G$.

$$\{l \leftarrow e, \neg e\} \models \neg e$$
$$\{l \leftarrow e, \neg e\} \not\models l$$
$$\{l \leftarrow e, \neg e\} \not\models \neg l$$
Human Reasoning – A Classical Logic Approach

▶ Recall the examples:

\[
\begin{align*}
\{l \leftarrow e, e\} & \models l \quad \text{modus ponens} \\
\{l \leftarrow e, e, l \leftarrow t\} & \models l \quad \text{classical logic is monoton} \\
\{l \leftarrow e, e, l \leftarrow o\} & \models l \quad \text{upps, humans don’t do this}
\end{align*}
\]

\[
\begin{align*}
\{l \leftarrow e, \neg e\} & \not\models \neg l \quad \text{denial of antecedent} \\
\{l \leftarrow e, \neg e, l \leftarrow t\} & \not\models \neg l \\
\{l \leftarrow e, \neg e, l \leftarrow o\} & \not\models \neg l
\end{align*}
\]

▶ Conclusion classical logic is inadequate.

▷ Often mistakenly generalized to “logic is inadequate”.
Human Reasoning – A Computational Logic Approach

Goal find a logic which adequately models human reasoning.

Solution I propose the following:

- Logic programs under (weak) completion semantics
  - Non-monotonicity
- Reasoning towards an appropriate logical form
  - Logic programs
- Three-valued Lukasiewicz logic
  - Least models
- An appropriate semantic operator
  - Least fixed points are least models
  - Least fixed points can be computed by iterating the operator
- Reasoning with respect to the least models
- A connectionist realization
Logic Programs

► A (logic) program is a finite set of clauses.
  ▶ A (program) clause is an expression of the form \( A \leftarrow B_1 \land \cdots \land B_n \), where \( n \geq 1 \), \( A \) is an atom, and each \( B_i, 1 \leq i \leq n \), is either a literal, \( \top \) or \( \bot \).
  ▶ \( A \) is called head and \( B_1 \land \cdots \land B_n \) body of the clause.
  ▶ A clause of the form \( A \leftarrow \top \) is called a positive fact.
  ▶ A clause of the form \( A \leftarrow \bot \) is called a negative fact.

\[
\{l \leftarrow e, e \leftarrow \top\} \\
\{l \leftarrow e, e \leftarrow \top, l \leftarrow t\} \\
\{l \leftarrow e, e \leftarrow \bot\}
\]

► Here I consider only propositional programs, but the approach extends to first-order programs.

► The language \( L \) underlying a program \( \mathcal{P} \) shall contain precisely the relation symbols occurring in \( \mathcal{P} \), and no others.
Program Completion

Let $\mathcal{P}$ be a program. Consider the following transformation:

1. All clauses with the same head $A \leftarrow Body_1$, $A \leftarrow Body_2$, \ldots are replaced by $A \leftarrow Body_1 \lor Body_2 \lor \ldots$.
2. If an atom $A$ is not the head of any clause in $\mathcal{P}$ then add $A \leftarrow \bot$.
3. All occurrences of $\leftarrow$ are replaced by $\leftrightarrow$.

The resulting set is called completion of $\mathcal{P}$ or $c\,\mathcal{P}$.
If step 2 is omitted then the resulting set is called weak completion of $\mathcal{P}$ or $wc\,\mathcal{P}$. 
Program Completion – Example 1

\[ \mathcal{P}_1 = \{ l \leftarrow e, e \leftarrow \top \} \]
\[ c \mathcal{P}_1 = \{ l \leftrightarrow e, e \leftrightarrow \top \} \]
\[ wc \mathcal{P}_1 = \{ l \leftrightarrow e, e \leftrightarrow \top \} \]

▶ The models of \( c \mathcal{P}_1 \) and \( wc \mathcal{P}_1 \)

▶ Hence, \( c \mathcal{P}_1 \models l \) and \( wc \mathcal{P}_1 \models l \)
Program Completion – Example 2

\[ P_2 = \{ l \leftarrow e, \ e \leftarrow \bot \} \]
\[ c(P_2) = \{ l \leftrightarrow e, \ e \leftrightarrow \bot \} \]
\[ wc(P_2) = \{ l \leftrightarrow e, \ e \leftrightarrow \bot \} \]

▶ The models of \( c P_2 \) and \( wc P_2 \)

▶ Hence, \( c P_2 \models \neg l \) and \( wc P_2 \models \neg l \)
▶ Remember, \( P_2 \nvdash \neg l \)
Program Completion – Example 3

\[ \mathcal{P}_3 = \{ l \leftarrow e, e \leftarrow \top, l \leftarrow t \} \]
\[ c\mathcal{P}_3 = \{ l \leftrightarrow e \lor t, e \leftrightarrow \top, t \leftrightarrow \bot \} \]
\[ wc\mathcal{P}_3 = \{ l \leftrightarrow e \lor t, e \leftrightarrow \top \} \]

- The models of \( c\mathcal{P}_3 \)
  \[ \{ e, l \} \]

- The models of \( wc\mathcal{P}_3 \)
  \[ \{ e, l \} \]
  \[ \{ e, l, t \} \]

- Hence, \( c\mathcal{P}_3 \models \neg t \) whereas \( wc\mathcal{P}_3 \not\models \neg t \) and \( wc\mathcal{P}_3 \not\models t \).
Monotonicity

- Let $\mathcal{F}$ and $\mathcal{F}'$ be sets of formulas and $G$ a formula.
  A logic is **monotonic** if the following holds:
  If $\mathcal{F} \models G$ then $\mathcal{F} \cup \mathcal{F}' \models G$.

- Classical logic is monotonic.

- A logic based on completion semantics is non-monotonic.

  - Consider
    \[
    \mathcal{P}_3 = \{ l \leftarrow e, e \leftarrow T, l \leftarrow t \} \\
    \mathcal{P}'_3 = \mathcal{P} \cup \{ t \leftarrow T \}
    \]

  - Then
    \[
    c_{\mathcal{P}_3} \models \neg t \\
    c_{\mathcal{P}'_3} \not\models \neg t
    \]
Reasoning Towards an Appropriate Logical Form

- Stenning, van Lambalgen 2008
- Represent conditionals as licences for conditionals.
  - If she has an essay to write then she will study late in the library.
    - She has an essay to write.
      \[ P_4 = \{ l \leftarrow e \land \neg ab, \ ab \leftarrow \bot, \ e \leftarrow \top \} \]
  - If she has an essay to write then she will study late in the library.
    - She has an essay to write.
      - If she has a textbook to read she will study late in the library.
      \[ P_5 = \{ l \leftarrow e \land \neg ab_1, \ ab_1 \leftarrow \bot, \ l \leftarrow t \land \neg ab_2, \ ab_2 \leftarrow \bot, \ e \leftarrow \top \} \]
  - Reason about additional premises.
    - If she has an essay to write then she will study late in the library.
      - She has an essay to write.
        - If the library stays open she will study late in the library.
        \[ P_6 = \{ l \leftarrow e \land \neg ab_1, \ ab_1 \leftarrow \neg o, \ l \leftarrow o \land \neg ab_2, \ ab_2 \leftarrow \neg e, \ e \leftarrow \top \} \]
Reasoning Towards an Appropriate Logical Form

» Denial of Antecedent (DA)

► If she has an essay to write then she will study late in the library. She does not have an essay to write.

►\[ P_7 = \{l \leftarrow e \land \neg ab, \; ab \leftarrow \bot, \; e \leftarrow \bot\} \]

► If she has an essay to write then she will study late in the library. She does not have an essay to write. If she has a textbook to read she will study late in the library.

►\[ P_8 = \{l \leftarrow e \land \neg ab_1, \; ab_1 \leftarrow \bot, \; l \leftarrow t \land \neg ab_2, \; ab_2 \leftarrow \bot, \; e \leftarrow \bot\} \]

► If she has an essay to write then she will study late in the library. She does not have an essay to write. If the library stays open she will study late in the library.

►\[ P_9 = \{l \leftarrow e \land \neg ab_1, \; ab_1 \leftarrow \neg o, \; l \leftarrow o \land \neg ab_2, \; ab_2 \leftarrow \neg e, \; e \leftarrow \bot\} \]
## Three-Valued Logics

<table>
<thead>
<tr>
<th></th>
<th>¬</th>
<th>∧</th>
<th>∨</th>
<th>←L</th>
<th>↔L</th>
<th>←K</th>
<th>↔K</th>
<th>↔C</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>⊤</td>
<td>⊤</td>
<td>⊤</td>
<td>⊤</td>
<td>⊤</td>
<td>⊤</td>
<td>⊤</td>
</tr>
<tr>
<td>T</td>
<td>⊥</td>
<td>⊥</td>
<td>⊤</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
</tr>
<tr>
<td>T</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>⊤</td>
<td>⊤</td>
<td>U</td>
<td>T</td>
<td>⊥</td>
</tr>
<tr>
<td>⊥</td>
<td>U</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
<td>⊥</td>
</tr>
<tr>
<td>⊥</td>
<td>U</td>
<td>⊤</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>T</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>T</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>T</td>
<td>U</td>
<td>U</td>
<td>⊥</td>
</tr>
<tr>
<td>U</td>
<td>⊥</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>T</td>
<td>U</td>
<td>U</td>
<td>⊥</td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>T</td>
<td>T</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>T</td>
<td>⊤</td>
<td>T</td>
<td>U</td>
<td>⊤</td>
</tr>
</tbody>
</table>

**Lukasiewicz (L) semantics 1920**

**Kleene (K) semantics 1952**

**Fitting (F) semantics 1985**

\[
(F \leftrightarrow G) \begin{cases} \equiv_{3L} \quad & (F \leftrightarrow G) \land (G \leftrightarrow F) \\
\neq_{3F} \quad & \end{cases}
\]
A (three-valued) interpretation is a mapping $\mathcal{L} \mapsto \{\top, \bot, U\}$ represented by $\langle I^\top, I^\bot \rangle$, where

- $I^\top$ contains all atoms which are mapped to $\top$,
- $I^\bot$ contains all atoms which are mapped to $\bot$,
- $I^\top \cap I^\bot = \emptyset$.
- All atoms which occur neither in $I^\top$ nor $I^\bot$ are mapped to $U$. 

Three-Valued Interpretations
Interpretations and Models

Let $\mathcal{I}$ denote the set of all interpretations.

Fitting 1985 $(\mathcal{I}, \subseteq)$ is a complete semi-lattice.

An interpretation $I$ is a model for a program $\mathcal{P}$, in symbols $I \models_3 \mathcal{P}$, iff $I(\mathcal{P}) = \top$.

\[
\begin{align*}
\langle \{p, q\}, \emptyset \rangle & \rightarrow \langle \{p\}, \emptyset \rangle & \rightarrow \langle \{q\}, \emptyset \rangle & \rightarrow \langle \emptyset, \{p\} \rangle \\
\langle \{p\}, \emptyset \rangle & \rightarrow \langle \{q\}, \emptyset \rangle & \rightarrow \langle \emptyset, \{q\} \rangle & \rightarrow \langle \emptyset, \{p\} \rangle
\end{align*}
\]
Logic Programs under Three-Valued $L$-Semantics

- We consider Lukasiewicz semantics.
- Let $\mathcal{P}$ be a logic program.
- **Theorem 1** The model intersection property holds for $\mathcal{P}$, i.e., $\cap\{I \mid I \models_{3L} \mathcal{P}\} \models_{3L} \mathcal{P}$.
  - If $\langle I^T, I^\perp \rangle \models_{3L} \mathcal{P}$, then $\langle I^T, \emptyset \rangle \models_{3L} \mathcal{P}$.
  - If $\langle I_1^T, \emptyset \rangle \models_{3L} \mathcal{P}$ and $\langle I_2^T, \emptyset \rangle \models_{3L} \mathcal{P}$, then $\langle I_1^T \cap I_2^T, \emptyset \rangle \models_{3L} \mathcal{P}$.
- Let $\text{lm}_{3L} \mathcal{P}$ denote the least $L$-model of $\mathcal{P}$.
  \[
  \text{lm}_{3L} \{p \leftarrow q\} = \langle \emptyset, \emptyset \rangle
  \]

- **Observation** Theorem 1 does not hold under $F$-semantics:
  - $\langle\{p, q\}, \emptyset\rangle \models_{3F} \{p \leftarrow q\}$ but $\langle\emptyset, \emptyset\rangle \not\models_{3F} \{p \leftarrow q\}$
  - $\langle\emptyset, \{p, q\}\rangle \models_{3F} \{p \leftarrow q\}$
Weakly Completed Logic Programs under L-Semantics

- We consider Lukasiewicz semantics.
- Let $\mathcal{P}$ be a logic program.
- **Theorem 2** The model intersection property holds for $wc\ \mathcal{P}$ as well.
- **Theorem 3** If $I \models_{3L} wc\ \mathcal{P}$ then $I \models_{3L} \mathcal{P}$.
- **Observation** Theorem 3 does not hold under F-semantics.

$$\langle \emptyset, \emptyset \rangle \models_{3F} wc\ \{p \leftarrow q\} = \{p \leftrightarrow q\}, \text{ but } \langle \emptyset, \emptyset \rangle \not\models_{3F} \{p \leftarrow q\}$$
Reasoning with Respect to the Least Model of \( wc \mathcal{P} \)

- **Recall our examples**

\[
\begin{align*}
wc \mathcal{P}_4 &= \{ l \leftrightarrow e \land \neg ab, \ ab \leftrightarrow \bot, \ e \leftrightarrow \top \} \\
lm_{3L} wc \mathcal{P}_4 &= \langle \{ e, l \}, \{ ab \} \rangle \leadsto \lm_{3L} wc \mathcal{P}_4(l) = \top \\
wcc \mathcal{P}_5 &= \{ l \leftrightarrow (e \land \neg ab_1) \lor (t \land \neg ab_2), \ ab_1 \leftrightarrow \bot, \ ab_2 \leftrightarrow \bot, \ e \leftrightarrow \top \} \\
lm_{3L} wc \mathcal{P}_5 &= \langle \{ e, l \}, \{ ab_1, ab_2 \} \rangle \leadsto \lm_{3L} wc \mathcal{P}_5(l) = \top \\
wcc \mathcal{P}_6 &= \{ l \leftrightarrow (e \land \neg ab_1) \lor (o \land \neg ab_2), \ ab_1 \leftrightarrow \neg o, \ ab_2 \leftrightarrow \neg e, \ e \leftrightarrow \top \} \\
lm_{3L} wc \mathcal{P}_6 &= \langle \{ e \}, \{ ab_2 \} \rangle \leadsto \lm_{3L} wc \mathcal{P}_6(l) = U \\
wcc \mathcal{P}_7 &= \{ l \leftrightarrow e \land \neg ab, \ ab \leftrightarrow \bot, \ e \leftrightarrow \bot \} \\
lm_{3L} wc \mathcal{P}_7 &= \langle \emptyset, \{ e, l, ab \} \rangle \leadsto \lm_{3L} wc \mathcal{P}_7(l) = \bot \\
wcc \mathcal{P}_8 &= \{ l \leftrightarrow (e \land \neg ab_1) \lor (t \land \neg ab_2), \ ab_1 \leftrightarrow \bot, \ ab_2 \leftrightarrow \bot, \ e \leftrightarrow \bot \} \\
lm_{3L} wc \mathcal{P}_8 &= \langle \emptyset, \{ e, ab_1, ab_2 \} \rangle \leadsto \lm_{3L} wc \mathcal{P}_8(l) = U \\
wcc \mathcal{P}_9 &= \{ l \leftrightarrow (e \land \neg ab_1) \lor (o \land \neg ab_2), \ ab_1 \leftrightarrow \neg o, \ ab_2 \leftrightarrow \neg e, \ e \leftrightarrow \bot \} \\
lm_{3L} wc \mathcal{P}_9 &= \langle \{ ab_2 \}, \{ e, l \} \rangle \leadsto \lm_{3L} wc \mathcal{P}_9(l) = \bot
\end{align*}
\]

- **This logic is adequate!**
Completion versus Weak Completion

- Recall $\mathcal{P}_8$

\[
\begin{align*}
\text{wc } \mathcal{P}_8 &= \{l \leftrightarrow (e \land \neg ab_1) \lor (t \land \neg ab_2), \ ab_1 \leftrightarrow \bot, \ ab_2 \leftrightarrow \bot, \ e \leftrightarrow \bot\} \\
\text{lm}_3 \text{L wc } \mathcal{P}_8 &= \langle \emptyset, \{e, ab_1, ab_2\} \rangle \leadsto \text{lm}_3 \text{L wc } \mathcal{P}_8(l) = U \\
\text{c } \mathcal{P}_8 &= \{l \leftrightarrow (e \land \neg ab_1) \lor (t \land \neg ab_2), \ ab_1 \leftrightarrow \bot, \ ab_2 \leftrightarrow \bot, \ e \leftrightarrow \bot, \ t \leftrightarrow \bot\} \\
\text{lm}_3 \text{L c } \mathcal{P}_8 &= \langle \emptyset, \{e, t, l, ab_1, ab_2\} \rangle \leadsto \text{lm}_3 \text{L c } \mathcal{P}_8(l) = \bot
\end{align*}
\]

- A logic using completion instead of weak completion is not adequate.
Computing the Least Models of Weakly Completed Programs

► How can we compute the least models of weakly completed programs?
► A first candidate:
  Fitting’s immediate consequence operator $\Phi_{F,P}(I) = \langle J^\top, J^\perp \rangle$, where

  $J^\top = \{ A \mid \text{there exists } A \leftarrow \text{Body} \in P \text{ with } I(\text{Body}) = \top \}$ and
  $J^\perp = \{ A \mid \text{for all } A \leftarrow \text{Body} \in P \text{ we find } I(\text{Body}) = \bot \}.$

► Uses F-semantics.
► Some well-known results mostly due to Fitting 1985:
  (1) $\Phi_{F,P}$ is monotone on $(I, \subseteq)$.
  (2) $\Phi_{F,P}$ is continuous
      and, hence, admits a least fixed point denoted by $\lf p \Phi_{F,P}$.
  (3) $\lf p \Phi_{F,P}$ can be computed by iterating $\Phi_{F,P}$ on $\langle \emptyset, \emptyset \rangle$.
  (3) The least F-model of $c P$ is the least fixed point of $\Phi_{F,P}$.

► Inadequate for human reasoning.
The Stenning and van Lambalgen Operator

- Stenning and van Lambalgen’s operator $\Phi_{\text{SvL}, \mathcal{P}}(I) = \langle J^\top, J^\bot \rangle$, where

  $$J^\top = \{ A \mid \text{there exists } A \leftarrow \text{Body} \in \mathcal{P} \text{ with } I(\text{Body}) = \top \}$$

  $$J^\bot = \{ A \mid \text{there exists } A \leftarrow \text{Body} \in \mathcal{P} \text{ and for all } A \leftarrow \text{Body} \in \mathcal{P} \text{ we find } I(\text{Body}) = \bot \}.$$  

- Theorem 4
  1. $\Phi_{\text{SvL}, \mathcal{P}}$ is monotone on $(\mathcal{I}, \subseteq)$.
  2. $\Phi_{\text{SvL}, \mathcal{P}}$ is continuous and, hence, admits a least fixed point denoted by $\text{lfp } \Phi_{\text{SvL}, \mathcal{P}}$.
  3. $\text{lfp } \Phi_{\text{SvL}, \mathcal{P}}$ can be computed by iterating $\Phi_{\text{SvL}, \mathcal{P}}$ on $\langle \emptyset, \emptyset \rangle$.
  4. $\text{lfp } \text{lm}_{3L \text{ wc } \mathcal{P}} = \text{lfp } \Phi_{\text{SvL}, \mathcal{P}}$. 

Steffen Hölldobler
Human Reasoning, Logic Programs and the CORE Method
Computing Least Fixed Points

- Recall some of our examples

\[ \mathcal{P}_4 = \{ l \leftarrow e \land \neg ab, \ ab \leftarrow \perp, \ e \leftarrow T \} \]

\[ \text{wc}\ \mathcal{P}_4 = \{ l \leftarrow e \land \neg ab, \ ab \leftarrow \perp, \ e \leftarrow T \} \]

\[ \Phi_{\text{SVL}, \mathcal{P}_4}(\langle \emptyset, \emptyset \rangle) = \langle \{ e \}, \{ ab \} \rangle \]

\[ \Phi_{\text{SVL}, \mathcal{P}_4}(\langle \{ e \}, \{ ab \} \rangle) = \langle \{ e, l \}, \{ ab \} \rangle = \text{lfp } \Phi_{\text{SVL}, \mathcal{P}_4} = \text{lm}_{3L} \text{ wc } \mathcal{P}_4 \]

\[ \mathcal{P}_9 = \{ l \leftarrow e \land \neg ab_1, \ ab_1 \leftarrow \neg o, \]

\[ l \leftarrow o \land \neg ab_2, \ ab_2 \leftarrow \neg e, \ e \leftarrow \perp \} \]

\[ \text{wc}\ \mathcal{P}_9 = \{ l \leftarrow (e \land \neg ab_1) \lor (o \land \neg ab_2), \]

\[ ab_1 \leftarrow \neg o, \ ab_2 \leftarrow \neg e, \ e \leftarrow \perp \} \]

\[ \Phi_{\text{SVL}, \mathcal{P}_9}(\langle \emptyset, \emptyset \rangle) = \langle \emptyset, \{ e \} \rangle \]

\[ \Phi_{\text{SVL}, \mathcal{P}_9}(\langle \emptyset, \{ e \} \rangle) = \langle \{ ab_2 \}, \{ e \} \rangle \]

\[ \Phi_{\text{SVL}, \mathcal{P}_9}(\langle \{ ab_2 \}, \{ e \} \rangle) = \langle \{ ab_2 \}, \{ e, l \} \rangle = \text{lfp } \Phi_{\text{SVL}, \mathcal{P}_9} = \text{lm}_{3L} \text{ wc } \mathcal{P}_9 \]
Cores for Three-Valued Logic Programs

- Consider \( \{p \leftarrow q\} \).
- A translation algorithm translates programs into a core.
- Recurrent connections connect the output to the input layer.

Kalinke 1995, Seda, Lane 2004

\[ \Phi_F \]

\[
p \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad p
\]

\[
\neg p \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad \neg p
\]

\[
q \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad q
\]

\[
\neg q \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad \neg q
\]

new

\[ \Phi_{SvL} \]

\[
p \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad p
\]

\[
\neg p \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad \neg p
\]

\[
q \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad q
\]

\[
\neg q \quad \frac{\omega}{2} \quad \frac{\omega}{2} \quad \neg q
\]

\[
\bot \quad -0.5
\]

\[
\top \quad -0.5
\]
A CORE Method for Human Reasoning

- Consider three-layer feed forward networks of binary threshold units.
- The input as well as the output layer shall represent interpretations.
- **Theorem 5**
  For each program $\mathcal{P}$ there exists a feed-forward core computing $\Phi_{SvL, \mathcal{P}}$.
- Add recurrent connections between corresponding units in the output and the input layer.
- **Corollary 6**
  The recurrent network reaches a stable state representing $\text{lfp } \Phi_{SvL, \mathcal{P}}$ if initialized with $\langle \emptyset, \emptyset \rangle$. 
Human Reasoning – Modus Ponens

- If she has an essay to write, she will study late in the library. She has an essay to write.
- Byrne 1989 96% of subjects conclude that she will study late in the library.
- Stenning, van Lambalgen 2008 $\mathcal{P}_4 = \{ l \leftarrow e \land \neg ab, e \leftarrow \top, ab \leftarrow \bot \}$.

$lfp \Phi_{SVL}, \mathcal{P}_4 = \text{lim}_{n \to \infty} \text{wc} \mathcal{P}_4 = \langle \{ l, e \}, \{ ab \} \rangle$.
- From $\langle \{ l, e \}, \{ ab \} \rangle$ follows that she will study late in the library.
If she has an essay to write, she will study late in the library. She has an essay to write. If she has some textbooks to read, she will study late in the library.

Byrne 1989 96% of subjects conclude that she will study late in the library.

Stenning, van Lambalgen 2008
\[ \mathcal{P}_5 = \{ l \leftarrow e \land \neg ab_1, e \leftarrow \top, ab_1 \leftarrow \bot, l \leftarrow t \land \neg ab_2, ab_2 \leftarrow \bot \} \].

\[ \text{lfp } \Phi_{SVL}, \mathcal{P}_5 = \text{lm}_{3L} \text{ wc } \mathcal{P}_5 = \langle \{ e, l \}, \{ ab_1, ab_2 \} \rangle. \]

From \( \langle \{ e, l \}, \{ ab_1, ab_2 \} \rangle \) follows that she will study late in the library.
Human Reasoning – Additional Argument

- *If she has an essay to write, she will study late in the library. She has an essay to write. If the library stays open, she will study late in the library.*
- *Byrne 1989* 38% of subjects conclude that she will study late in the library.
- *Stenning, van Lambalgen 2008*

\[ P_6 = \{ l \leftarrow e \land \neg ab_1, e \leftarrow \top, l \leftarrow o \land \neg ab_2, ab_1 \leftarrow \neg o, ab_2 \leftarrow \neg e, \} \]

\[
\begin{array}{cccccccccccc}
e & \neg e & l & \neg l & ab_1 & \neg ab_1 & o & \neg o & ab_2 & \neg ab_2 \\
\omega/2 & \omega/2 & \omega/2 & 3\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
3\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\omega/2 & \omega/2 & \omega/2 & 3\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\omega/2 & \omega/2 & \omega/2 & 3\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\omega/2 & \omega/2 & \omega/2 & 3\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\end{array}
\]

\[
\begin{array}{cccccccccccc}
\omega/2 & \omega/2 & \omega/2 & 3\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 & \omega/2 \\
\end{array}
\]

- \( \text{lfp } \Phi_{SvL,P_6} = \text{lm}_{3L \text{ wc }} P_6 = \langle \{e\}, \{ab_2\} \rangle \).
- *From \( \langle \{e\}, \{ab_2\} \rangle \) follows that it is unknown whether she will study late in the library.*
Human Reasoning – Denial of Antecedent (DA)

- If she has an essay to write, she will study late in the library. She does not have an essay to write.
- Byrne 1989 46% of subjects conclude that she will not study late in the library.
- Stenning, van Lambalgen 2008 \( P_7 = \{ l \leftarrow e \land \neg ab, e \leftarrow \bot, ab \leftarrow \bot \} \).

\[
\text{\small \( \text{lfp} \Phi_{\text{SvL},P_7} = \text{lm}_3 \text{wc} P_7 = \langle \emptyset, \{ab, e, l\} \rangle \).}
\]

- From \( \langle \emptyset, \{ab, e, l\} \rangle \) follows that she will not study late in the library.
**Human Reasoning – Alternative Argument and DA**

- *If she has an essay to write, she will study late in the library. She does not have an essay to write. If she has textbooks to read, she will study late in the library.*

- *Byrne 1989* 4% of subjects conclude that Marian will not study late in the library.

- *Stenning, van Lambalgen 2008*
  \[ P_8 = \{ l \leftarrow e \land \neg ab_1, e \leftarrow \bot, ab_1 \leftarrow \bot, l \leftarrow t \land \neg ab_2, ab_2 \leftarrow \bot \} \].

\[ \lf p \Phi_{SvL, P_8} = \lim_{3L} wc P_8 = \langle \emptyset, \{ ab_1, ab_2, e \} \rangle. \]

- *From* \( \langle \emptyset, \{ ab_1, ab_2, e \} \rangle \) *follows that it is unknown whether she will study late in the library.*
Human Reasoning – Additional Argument and DA

- *If she has an essay to write, she will study late in the library. She does not have an essay to write. If the library is open, she will study late in the library.*

- *Byrne 1989* 63% of subjects conclude that she will not study late in the library.

- *Stenning, van Lambalgen 2008*

\[
P_9 = \{l \leftarrow e \land \neg ab_1, e \leftarrow \bot, l \leftarrow o \land \neg ab_2, ab_1 \leftarrow \neg o, ab_2 \leftarrow \neg e\}.
\]

- \(\text{lfp } \Phi_{SVL}, P_9 = \text{Im}_{3L} \text{wc } P_9 = \langle \{ab_2\}, \{e, l\} \rangle\).

- From \(\langle \{ab_2\}, \{e, l\} \rangle\) follows that she will not study late in the library.
Summary

- Under Lukasiewicz semantics we obtain

<table>
<thead>
<tr>
<th>Modus Ponens</th>
<th>Byrne 1989</th>
<th>Program</th>
<th>Im$_{3L}$ wc $P_i(l)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>l (96%)</td>
<td>$P_4$</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>Alternative Arguments</td>
<td>l (96%)</td>
<td>$P_5$</td>
<td>T</td>
</tr>
<tr>
<td>Additional Arguments</td>
<td>l (38%)</td>
<td>$P_6$</td>
<td>U</td>
</tr>
<tr>
<td>Modus Ponens and DA</td>
<td>$\neg l$ (46%)</td>
<td>$P_7$</td>
<td>$\perp$</td>
</tr>
<tr>
<td>Alternative Arguments and DA</td>
<td>$\neg l$ (4%)</td>
<td>$P_8$</td>
<td>U</td>
</tr>
<tr>
<td>Additional Arguments and DA</td>
<td>$\neg l$ (63%)</td>
<td>$P_9$</td>
<td>$\perp$</td>
</tr>
</tbody>
</table>

- The approach appears to be adequate.
- Fitting semantics or completion is inadequate.
Contraction Mappings

- Do we have to initialize the networks with the empty interpretation?

- **Banach’s Contraction Mapping Theorem**
  A contraction mapping $f$ on a complete metric space has a unique fixed point; the sequence $x, f(x), f(f(x)), \ldots$ converges to this fixed point, where $x$ is an arbitrary element from the metric space.

- Let $P$ be a program.
  A **level mapping** is a mapping $l$ from the set of atoms to the set of natural numbers. It is extended to negative atoms by defining $l(\neg A) = l(A)$ for each atom $A$.

- Let $\mathcal{I}$ be the set of all interpretations and $I, J \in \mathcal{I}$.

  $$d_l(I, J) = \begin{cases} \frac{1}{2^n} & \text{if } I \neq J \text{ and } I(A) = J(A) \neq \emptyset \text{ for all } A \text{ with } l(A) < n \text{ and } \text{I}(A) \neq J(A) \text{ or } I(A) = J(A) = \emptyset \text{ for some } A \text{ with } l(A) = n, \\ 0 & \text{otherwise.} \end{cases}$$

- **Proposition 7 (Kencana Ramli 2009)** $(\mathcal{I}, d_l)$ is a complete metric space.
Contraction Properties of the Semantic Operators

> Observation
If \( \mathcal{P} \) is an acceptable program then \( \Phi_{SVL,\mathcal{P}} \) is not necessarily a contraction.

- Consider \( \mathcal{P} = \{ r \land q \to p, \ r \land p \to q \} \).
- \( \mathcal{P} \) is acceptable.
- Both, \( \langle \emptyset, \emptyset \rangle \) and \( \langle \emptyset, \{ p, q \} \rangle \), are fixed points of \( \Phi_{SVL,\mathcal{P}} \).
- By Banach’s contraction mapping theorem \( \Phi_{SVL,\mathcal{P}} \) is not a contraction.

> Theorem 8 If \( \mathcal{P} \) is an acyclic program then \( \Phi_{SVL,\mathcal{P}} \) is a contraction.

- If \( \mathcal{P} \) is acyclic then we find a level mapping \( l \) such that for each \( L_1 \land \ldots \land L_n \to A \in \mathcal{P} \) we have \( l(L_i) < l(A) \).
- \( d_l(\Phi_{SVL,\mathcal{P}}(I), \Phi_{SVL,\mathcal{P}}(J)) \leq \frac{1}{2}d_l(I, J) \).

> Corollary 9 If \( \mathcal{P} \) is an acyclic program then \( \Phi_{SVL,\mathcal{P}} \) has a unique fixed point which can be reached by iterating \( \Phi_{SVL,\mathcal{P}} \) starting from some interpretation.
Positive versus Negative Information

- Consider $\mathcal{P} = \{p \leftarrow \top, p \leftarrow \bot\}$.
- $\text{Im}_{3L \text{ wc}} \mathcal{P} = \langle \{p\}, \emptyset \rangle$.

Positive information dominates negative information.

However, we may take integrity constraints into account, i.e. expressions of the form $L_1 \land \ldots \land L_n \rightarrow \bot$. 
Long Term versus Short Term Memory

Consider $\mathcal{P} = \{p \leftarrow \bot\}$.

If unit $p$ in the output layer is clamped, then the input and output layers do not represent interpretations any more.

Adding a connection between the units $p$ and $\neg p$ in the output layer with weight $-\omega$ yields the desired effect.
Human Reasoning – Affirmation of the Consequent (AC)

▶ If she has an essay to write then she will study late in the library. She will study late in the library.

▷ Byrne 1989 53% of subjects conclude that she has an essay to write.

▶ If she has an essay to write then she will study late in the library. She will study late in the library. If she has a textbook to read she will study late in the library.

▷ Byrne 1989 16% of subjects conclude that she has an essay to write.

▶ If she has an essay to write then she will study late in the library. She will study late in the library. If the library stays open she will study late in the library.

▷ Byrne 1989 55% of subjects conclude that she has an essay to write.
Human Reasoning – Modus Tollens

- If she has an essay to write then she will study late in the library. She will not study late in the library.
  - Byrne 1989 69% of subjects conclude that she does not have an essay to write.

- If she has an essay to write then she will study late in the library. She will not study late in the library. If she has a textbook to read she will study late in the library.
  - Byrne 1989 69% of subjects conclude that she does not have an essay to write.

- If she has an essay to write then she will study late in the library. She will not study late in the library. If the library stays open she will study late in the library.
  - Byrne 1989 44% of subjects conclude that she does not have an essay to write.
Abductive Frameworks and Observations

▶ Remember

▷ Let $\mathcal{K} \subseteq \mathcal{L}$ be a set of formulas called the knowledge base, $\mathcal{A} \subseteq \mathcal{L}$ be a set of formulas called abducibles, and $\models \subseteq 2\mathcal{L} \times \mathcal{L}$ a logical consequence relation. $\langle \mathcal{K}, \mathcal{A}, \models \rangle$ is called abductive framework.

▷ An observation $\mathcal{O}$ is a subset of the language $\mathcal{L}$.

▶ Here

▷ $\mathcal{K}$ is a logic program $\mathcal{P}$ (with negative facts).

▷ $\mathcal{L}$ is the language underlying $\mathcal{P}$.

▷ $\mathcal{R}^D_\mathcal{P} = \{ A \in \mathcal{R}_\mathcal{P} \mid A \leftarrow Body \in \mathcal{P} \}$ is the set of defined predicates in $\mathcal{P}$.

▷ $\mathcal{R}^U_\mathcal{P} = \mathcal{R}_\mathcal{P} \setminus \mathcal{R}^D_\mathcal{P}$ is the set of undefined predicates in $\mathcal{P}$.

▷ $\mathcal{A}$ is the set $\{ A \leftarrow \top \mid A \in \mathcal{R}^U_\mathcal{P} \} \cup \{ A \leftarrow \bot \mid A \in \mathcal{R}^U_\mathcal{P} \}$.

▷ $\models$ is $\models^{\text{im\,wc}}_{3\mathcal{L}}$, where $\mathcal{P} \models^{\text{im\,wc}}_{3\mathcal{L}} F$ iff $\text{Im}_{3\mathcal{L}} \text{wc } \mathcal{P}(F) = \top$.

▷ Observations are usually sets containing a single literal, in which case we simply write $\mathcal{O} = L$ instead of $\mathcal{O} = \{ L \}$. 
Explanations

- Let $\mathcal{L}$ be a language.
- Let $\langle \mathcal{K}, \mathcal{A}, \models \rangle$ be an abductive framework and $\mathcal{O}$ an observation.
- $\mathcal{O}$ is explained by $\mathcal{E}$ (or $\mathcal{E}$ is an explanation for $\mathcal{O}$) if
  $\triangleright \mathcal{E} \subseteq \mathcal{A}$,
  $\triangleright \mathcal{K} \cup \mathcal{E}$ is satisfiable,
  $\triangleright \mathcal{K} \cup \mathcal{E} \models L$ for each $L \in \mathcal{O}$.
- An explanation $\mathcal{E}$ for $\mathcal{O}$ is said to be a minimal iff there is no explanation $\mathcal{E}' \subset \mathcal{E}$ for $\mathcal{O}$. 
Reasoning

Let \( \langle \mathcal{P}, \mathcal{A}, \models_{3L}^{\text{im wc}} \rangle \) be an abductive framework, where

- \( \mathcal{P} \) is a logic program and
- \( \mathcal{A} = \{ A \leftarrow \top | A \in \mathcal{R}_P^U \} \cup \{ A \leftarrow \bot | A \in \mathcal{R}_P^U \} \) the set of abducibles

Let \( \mathcal{O} \) be an observation and \( F \) a formula in the language underlying \( \mathcal{P} \).

- \( F \) follows sceptically by abduction from \( \mathcal{P} \) and \( \mathcal{O} \) (in symbols \( \mathcal{P}, \mathcal{O} \models^s_A F \)) iff \( \mathcal{O} \) can be explained and for all minimal explanations \( \mathcal{E} \) we find \( \mathcal{P} \cup \mathcal{E} \models_{3L}^{\text{im wc}} F \).

- \( F \) follows credulously by abduction from \( \mathcal{P} \) and \( \mathcal{O} \) (in symbols \( \mathcal{P}, \mathcal{O} \models^c_A F \)) iff there exists a minimal explanation \( \mathcal{E} \) such that \( \mathcal{P} \cup \mathcal{E} \models_{3L}^{\text{im wc}} F \).
Human Reasoning – Modus Ponens and AC

▶ If she has an essay to write then she will study late in the library.
She will study late in the library.

▷ Byrne 1989 53% of subjects conlude that she has an essay to write.

▶ We obtain

\[ P_{10} = \{ l \leftarrow e \land \neg ab, ab \leftarrow \bot \} \]
\[ A = \{ e \leftarrow \top, e \leftarrow \bot \} \]
\[ \mathcal{O} = l \]

▶ Thus

\[
\text{Im}_3 \text{wc } P_{10} = \langle \emptyset, \{ ab \} \rangle \\
\text{Im}_3 \text{wc } (P_{10} \cup \{ e \leftarrow \top \}) = \langle \{ e, l \}, \{ ab \} \rangle \\
\text{Im}_3 \text{wc } (P_{10} \cup \{ e \leftarrow \bot \}) = \langle \emptyset, \{ e, l, ab \} \rangle
\]

▶ Hence, \( \{ e \leftarrow \top \} \) is the only minimal explanation and \( P_{10}, \mathcal{O} \models^s_A e \).
Human Reasoning – Alternative Arguments and AC

If she has an essay to write then she will study late in the library.
She will study late in the library.
If she has a textbook to read she will study late in the library.

Byrne 1989 16% of subjects conclude that she has an essay to write.

We obtain

\[
P_{11} = \{l \leftarrow e \land \neg ab_1, ab_1 \leftarrow \bot, l \leftarrow t \land \neg ab_2, ab_2 \leftarrow \bot\}
\]
\[
A = \{e \leftarrow \top, e \leftarrow \bot, t \leftarrow \top, e \leftarrow \bot\}
\]
\[
\mathcal{O} = l
\]

Thus

\[
\text{Im}_{3L \text{wc}} P_{11} = \langle \emptyset, \{ab_1, ab_2\}\rangle
\]
\[
\text{Im}_{3L \text{wc}} (P_{11} \cup \{e \leftarrow \top\}) = \langle \{e, l\}, \{ab_1, ab_2\}\rangle
\]
\[
\text{Im}_{3L \text{wc}} (P_{11} \cup \{e \leftarrow \bot\}) = \langle \emptyset, \{e, ab_1, ab_2\}\rangle
\]
\[
\text{Im}_{3L \text{wc}} (P_{11} \cup \{t \leftarrow \top\}) = \langle \{t, l\}, \{ab_1, ab_2\}\rangle
\]
\[
\text{Im}_{3L \text{wc}} (P_{11} \cup \{t \leftarrow \bot\}) = \langle \emptyset, \{t, ab_1, ab_2\}\rangle
\]

Hence, \{e \leftarrow \top\} and \{t \leftarrow \top\} are minimal explanations and \(P_{11}, \mathcal{O} \not\models_A e\).
Human Reasoning – Additional Arguments and AC

- If she has an essay to write then she will study late in the library. She will study late in the library. If the library stays open she will study late in the library.

- Byrne 1989 55% of subjects conclude that she has an essay to write.

- We obtain

\[
P_{12} = \{l ← e \land \neg a_1, a_1 ← \neg o, l ← o \land \neg a_2, a_2 ← \neg e\} \\
A = \{e ← \top, e ← \bot, o ← \top, o ← \bot\} \\
O = l
\]

- Thus

\[
\text{lm}_3 L \text{wc } P_{12} = \langle \emptyset, \emptyset \rangle \\
\text{lm}_3 L \text{wc } (P_{12} \cup \{e ← \top\}) = \langle \{e\}, \{a_2\} \rangle \\
\text{lm}_3 L \text{wc } (P_{12} \cup \{e ← \bot\}) = \langle \{a_2\}, \{e, l\} \rangle \\
\text{lm}_3 L \text{wc } (P_{12} \cup \{o ← \top\}) = \langle \{t\}, \{a_1\} \rangle \\
\text{lm}_3 L \text{wc } (P_{12} \cup \{o ← \bot\}) = \langle \{a_1\}, \{o, l\} \rangle \\
\text{lm}_3 L \text{wc } (P_{12} \cup \{e ← \top, o ← \top\}) = \langle \{t, e, l\}, \{a_1, a_2\} \rangle
\]

- \(\{e ← \top, o ← \top\}\) is the only minimal explanation and \(P_{12}, O \models_A e\).
Human Reasoning – Modus Tollens (MT)

- If she has an essay to write then she will study late in the library. She will not study late in the library.
- **Byrne 1989** 69% of subjects conclude that she does not have an essay to write.

- We obtain

\[
P_{10} = \{l \leftarrow e \land \neg ab, ab \leftarrow \bot\}
\]
\[
A = \{e \leftarrow \top, e \leftarrow \bot\}
\]
\[
O = \neg l
\]

- Thus

\[
\text{Im}_{3L} wc P_{10} = \langle \emptyset, \{ab\} \rangle
\]
\[
\text{Im}_{3L} wc (P_{10} \cup \{e \leftarrow \top\}) = \langle \{e, l\}, \{ab\} \rangle
\]
\[
\text{Im}_{3L} wc (P_{10} \cup \{e \leftarrow \bot\}) = \langle \emptyset, \{e, l, ab\} \rangle
\]

- Hence, \(\{e \leftarrow \bot\}\) is the only minimal explanation and \(P_{10}, O \models_A \neg e\).
If she has an essay to write then she will study late in the library. She will not study late in the library. If she has a textbook to read she will study late in the library.

Byrne 1989 69% of subjects conclude that she does not have an essay to write.

We obtain

\[ P_{11} = \{l \leftarrow e \land \neg a b_1, a b_1 \leftarrow \bot, l \leftarrow t \land \neg a b_2, a b_2 \leftarrow \bot\} \]
\[ A = \{e \leftarrow \top, e \leftarrow \bot, t \leftarrow \top, e \leftarrow \bot\} \]
\[ O = \neg l \]

Thus

\[ \text{Im}_{3L \text{ wc}} P_{11} = \langle \emptyset, \{a b_1, a b_2\} \rangle \]
\[ \text{Im}_{3L \text{ wc}} (P_{11} \cup \{e \leftarrow \top\}) = \langle \{e, l\}, \{a b_1, a b_2\} \rangle \]
\[ \text{Im}_{3L \text{ wc}} (P_{11} \cup \{e \leftarrow \bot\}) = \langle \emptyset, \{e, a b_1, a b_2\} \rangle \]
\[ \text{Im}_{3L \text{ wc}} (P_{11} \cup \{t \leftarrow \top\}) = \langle \{t, l\}, \{a b_1, a b_2\} \rangle \]
\[ \text{Im}_{3L \text{ wc}} (P_{11} \cup \{t \leftarrow \bot\}) = \langle \emptyset, \{t, a b_1, a b_2\} \rangle \]
\[ \text{Im}_{3L \text{ wc}} (P_{11} \cup \{e \leftarrow \bot, t \leftarrow \bot\}) = \langle \emptyset, \{e, l, t, a b_1, a b_2\} \rangle \]

\{e \leftarrow \bot, t \leftarrow \bot\} is the only minimal explanation and \( P_{11}, O \models_s \neg e \).
Human Reasoning – Additional Arguments and MT

If she has an essay to write then she will study late in the library.
She will not study late in the library.
If the library stays open she will study late in the library.

Byrne 1989 44% of subjects conclude that she does not have an essay to write.

We obtain

\[
\mathcal{P}_{12} = \{ l \leftarrow e \land \neg ab_1, ab_1 \leftarrow \neg o, l \leftarrow o \land \neg ab_2, ab_2 \leftarrow \neg e \} \\
\mathcal{A} = \{ e \leftarrow \top, e \leftarrow \bot, o \leftarrow \top, o \leftarrow \bot \} \\
\mathcal{O} = \neg l
\]

Thus

\[
\text{lm}_{3L \text{ wc}} \mathcal{P}_{12} = \langle \emptyset, \emptyset \rangle \\
\text{lm}_{3L \text{ wc}} (\mathcal{P}_{12} \cup \{ e \leftarrow \top \}) = \langle \{ e \}, \{ ab_2 \} \rangle \\
\text{lm}_{3L \text{ wc}} (\mathcal{P}_{12} \cup \{ e \leftarrow \bot \}) = \langle \{ ab_2 \}, \{ e, l \} \rangle \\
\text{lm}_{3L \text{ wc}} (\mathcal{P}_{12} \cup \{ o \leftarrow \top \}) = \langle \{ t \}, \{ ab_1 \} \rangle \\
\text{lm}_{3L \text{ wc}} (\mathcal{P}_{12} \cup \{ o \leftarrow \bot \}) = \langle \{ ab_1 \}, \{ o, l \} \rangle
\]

Hence, \{ e \leftarrow \bot \} and \{ o \leftarrow \bot \} are minimal explanations and \( \mathcal{P}_{12}, \mathcal{O} \not\models^s_A \neg e \).
Weak Completion is Needed

- Reconsider the case modus ponens with positive observation, i.e.
  \[ \mathcal{P}_{10} = \{ l \leftarrow e \land \neg ab, ab \leftarrow \bot \} \]
  \[ \mathcal{A} = \{ e \leftarrow \top, e \leftarrow \bot \} \]
  and
  \[ \mathcal{O} = l. \]

- Now consider \( \langle \mathcal{P}_{10}, \mathcal{A}, \models_{3L} \rangle \) instead of \( \langle \mathcal{P}_{10}, \mathcal{A}, \models_{3L}^{lm wc} \rangle \)
  \[ \mathcal{P}_{10} \not\models_{3L} l \]
  \[ \mathcal{P}_{10} \cup \{ e \leftarrow \top \} \not\models_{3L} l \) (because \( ab \) can be mapped to \( \top \)).
  \[ \mathcal{P}_{10} \cup \{ e \leftarrow \bot \} \not\models_{3L} l \]
  \[ \mathcal{P}_{10} \cup \mathcal{A} \not\models_{3L} l \]

Hence, the observation can not be explained at all (in contrast to Byrne 1989).
Completion is Insufficient

- Reconsider the case modus ponens with positive observation, i.e.
  \[ \Delta P_{10} = \{ l \leftrightarrow e \land \neg ab, ab \leftrightarrow \bot \}, \]
  \[ \Delta A = \{ e \leftrightarrow \top, e \leftrightarrow \bot \}, \]
  and
  \[ \Delta O = l. \]

- Now consider \( \langle P_{10}, A, \models^c_{3L} \rangle \) instead of \( \langle P_{10}, A, \models^{lm}_{3L} wc \rangle \),
  where \( P \models^c_{3L} F \) iff \( F \) holds in all models for \( c P \).
  \[ c P_{10} = \{ l \leftrightarrow e \land \neg ab, ab \leftrightarrow \bot, e \leftrightarrow \bot \} \models_{3L} \neg l \]
  \[ c P_{10} \models_{3L} \neg e \]
  The observation is inconsistent with the knowledge base and, thus, cannot be explained at all (in contrast to Byrne 1989).
Explanations must be Completed as well

► Reconsider the case modus ponens with positive observation, i.e.

\[ \Delta P_{10} = \{ l \leftarrow e \land \neg a b, a b \leftarrow \bot \}, \]
\[ \Delta A = \{ e \leftarrow \top, e \leftarrow \bot \}, \text{ and} \]
\[ \Delta O = l. \]

► Now (weakly) complete only the program, but not the explanations.

\[ \Delta wc P_{10} = \{ l \leftrightarrow e \land \neg a b, a b \leftrightarrow \bot \}. \]
\[ \Delta wc P_{10} \not\models_{3L} \neg l \]
\[ \Delta wc (P_{10}) \cup \{ e \leftarrow \top \} \not\models_{3L} \neg l \]
\[ \Delta wc (P_{10}) \cup \{ e \leftarrow \bot \} \not\models_{3L} \neg l \text{ (because } e \text{ can be mapped to } \top). \]
\[ \Delta wc (P_{10}) \cup A \not\models_{3L} \neg l \]

Hence, the observation can not be explained (in contrast with Byrne 1989).
Sceptical versus Credulous Reasoning

- Reconsider the case alternative arguments with positive observation, i.e.,
  \[ \phi_{11} = \{ l \leftarrow e \land \neg ab_1, ab_1 \leftarrow \bot, l \leftarrow t \land \neg ab_2, ab_2 \leftarrow \bot \}, \]
  \[ \psi = \{ e \leftarrow \top, e \leftarrow \bot \}, \text{ and} \]
  \[ \psi = l. \]

- Now consider \( \langle \phi_{11}, \psi, \models_{3L}^{\text{im wc}} \rangle \) and reason credulously:
  \[ \text{There are two minimal explanations, viz.} \ \{ e \leftarrow \top \} \text{ and} \ \{ e \leftarrow \bot \}. \]
  \[ \text{Hence,} \ \phi_{11}, l \not\models_{A}^{s} e, \text{ but} \ \phi_{11}, l \models_{A}^{c} e. \]

Credulous reasoning is inconsistent with Byrne 1989.
Summary

- Let \( P_{10} = \{ l \leftarrow e \land \neg ab, ab \leftarrow \bot \} \)
  
  \( P_{11} = \{ l \leftarrow e \land \neg ab_1, ab_1 \leftarrow \bot, l \leftarrow t \land \neg ab_2, ab_2 \leftarrow \bot \} \)
  
  \( P_{12} = \{ l \leftarrow e \land \neg ab_1, ab_1 \leftarrow \neg o, l \leftarrow o \land \neg ab_2, ab_2 \leftarrow \neg e \} \)

- We obtain

<table>
<thead>
<tr>
<th></th>
<th>Byrne 1989</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{10}, l \models_A e )</td>
<td>( e(53%) )</td>
</tr>
<tr>
<td>( P_{11}, l \not\models_A e )</td>
<td>( e(16%) )</td>
</tr>
<tr>
<td>( P_{12}, l \models_A e )</td>
<td>( e(55%) )</td>
</tr>
<tr>
<td>( P_{10}, \neg l \models_A \neg e )</td>
<td>( \neg e(69%) )</td>
</tr>
<tr>
<td>( P_{11}, \neg l \not\models_A \neg e )</td>
<td>( \neg e(69%) )</td>
</tr>
<tr>
<td>( P_{12}, \neg l \not\models_A e )</td>
<td>( \neg e(44%) )</td>
</tr>
</tbody>
</table>
Summary

▶ Logic appears to be adequate for human reasoning if
  ▶ weak completion,
  ▶ Lukasiewicz semantics,
  ▶ the Stenning and van Lambalgen semantic operator, and
  ▶ abduction are used

▶ Human Reasoning is modeled by
  ▶ reasoning towards an appropriate logic program $\mathcal{P}$ and, thereafter,
  ▶ reasoning with respect to the least L-model of the weak completion of $\mathcal{P}$.

▶ This approach matches data from studies in human reasoning.
▶ There is a connectionist encoding.
Discussion

- Stenning, van Lambalgen 2008 propose spreading-activation networks like KBANN (Towell, Shavlik 1993) with two units for each propositional letter and an inhibitory link between them.

- Logical threshold units can be replaced by bipolar sigmoidal ones following d’Avila Garcez, Zaverucha, Carbalho 1997.

  ▶ Networks can be trained by backpropagation,
  ▶ but backpropagation is not neurally plausible.
Some Open Problems (1)

- **Negation**
  - How is negation treated in human reasoning?

- **Errors**
  - How can frequently made errors be explained in the proposed approach?

- **Lukasievicz logic**
  - In a Lukasievicz logic the semantic deduction theorem does not hold.
  - Is this adequate with respect to human reasoning?

- **Completion**
  - Under which conditions is human reasoning adequately modeled by completion and/or weak completion?
Some Open Problems (2)

► Contractions
  ▶ Do humans exhibit a behavior which can be adequately modeled by contractional semantic operators?
  ▶ Can we generate appropriate level mappings by studying the behavior of humans?

► Explanations
  ▶ Do humans consider minimal explanations?
  ▶ In which order are (minimal) explanations generated by humans if there are several?
  ▶ Does attention play a role in the selection of (minimal) explanations?

► Stable coalitions
  ▶ Do stable coalitions occur in human reasoning?
  ▶ How are they deactivated?
Some Open Problems (3)

- **Reasoning**
  - Do humans reason sceptically or credulously?
  - How does a connectionist realization of sceptical reasoning looks like?

- **Theory revision**
  - How is theory revision modeled in human reasoning?

- **Relation to other semantics**
  - What is the relation between the proposed approach and well-founded and/or stable and/or projection semantics?