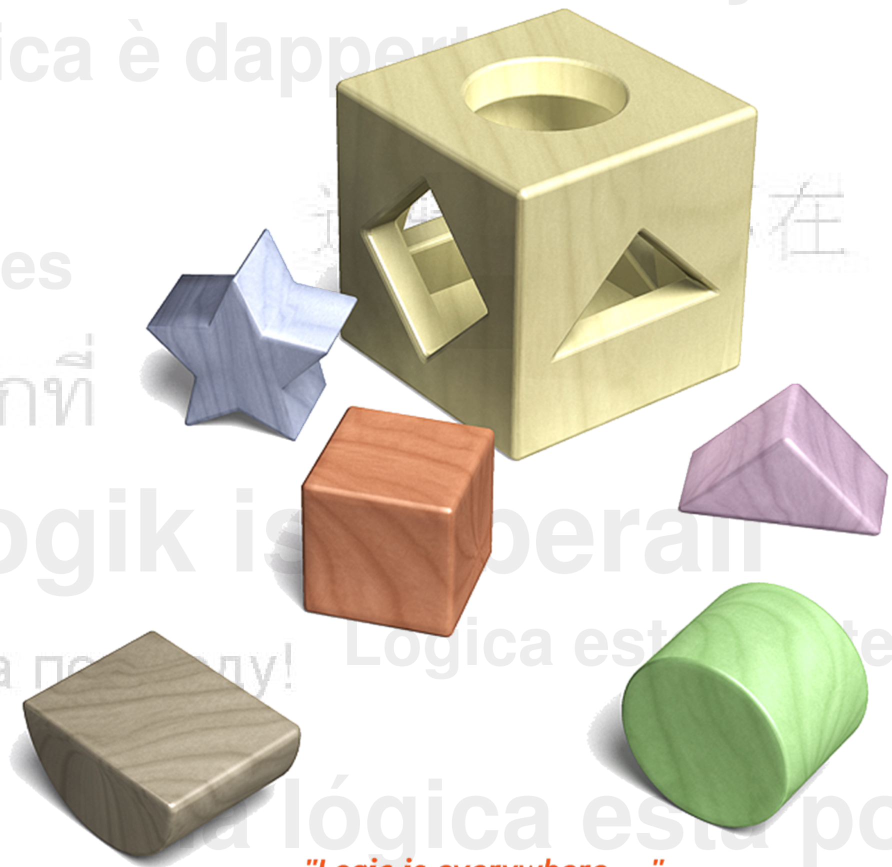
Computational Logic, Connectionist Systems and Human Reasoning

Steffen Hölldobler
Carroline Dewi Puspa Kencana Ramli
International Center for Computational Logic
Technische Universität Dresden
Germany

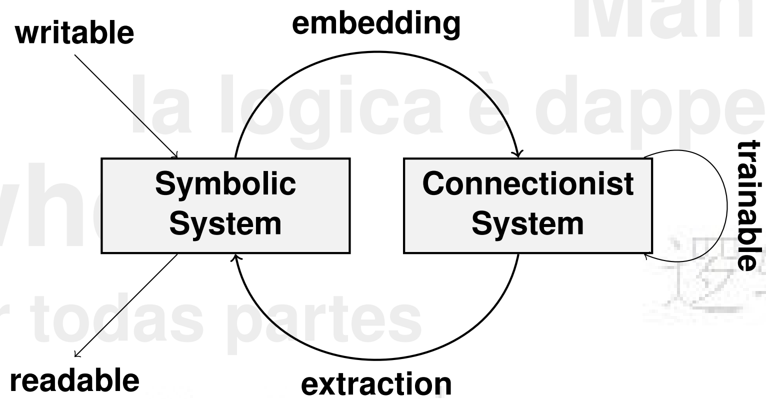
- **Motivation**
- **Three-Valued Logic Programs**
- **The Core Method**
- **Human Reasoning**



"Logic is everywhere ..."

Motivation

- ▶ **The Neural Symbolic Cycle**



- ▶ **The Core Method**
  Connectionist model generation using recurrent networks with feed-forward core.
- ▶ **A Challenge Problem**
  How does the core method relate to model-based reasoning approaches in cognitive science?
  - ▷ **“I can answer this one”**
    Michiel van Lambalgen at the ICCL summer school 2008 on Computational Logic and Cognitive Science.

Three-Valued Logics

		$\neg$	$\wedge$	$\vee$	$\leftarrow_K$	$\leftrightarrow_K$	$\leftarrow_N$	$\leftrightarrow_C$	$\leftarrow_Ł$	$\leftrightarrow_Ł$
$\top$	$\top$	$\bot$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$	$\bot$	$\top$	$\top$	$\bot$	$\bot$	$\bot$	$\top$	$\bot$
$\top$	U	$\bot$	U	$\top$	$\top$	U	$\bot$	$\bot$	$\top$	U
$\bot$	$\top$	$\top$	$\bot$	$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$
$\bot$	$\bot$	$\top$	$\bot$	$\bot$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$
$\bot$	U	$\top$	$\bot$	U	U	U	$\bot$	$\bot$	U	U
U	$\top$	U	U	$\top$	U	U	$\top$	$\bot$	U	U
U	$\bot$	U	$\bot$	U	$\top$	U	$\top$	$\bot$	$\top$	U
U	U	U	U	U	U	U	$\top$	$\top$	$\top$	$\top$

Łukasiewicz 1920	$\neg$	$\wedge$	$\vee$	$\leftarrow_Ł$	$\leftrightarrow_Ł$
Kleene 1952	$\neg$	$\wedge$	$\vee$	$\leftarrow_K$	$\leftrightarrow_K$
Fitting 1985	$\neg$	$\wedge$	$\vee$	$\leftarrow_K$	$\leftrightarrow_C$
Naish 2006	$\neg$	$\wedge$	$\vee$	$\leftarrow_N$	

$$(F \leftrightarrow G) \left\{ \begin{array}{c} \equiv_Ł \\ \not\equiv_F \end{array} \right\} (F \leftarrow G) \wedge (G \leftarrow F)$$

Logic Programs

- A **(logic) program** is a finite set of clauses.
  - A **(program) clause** is an expression of the form $A \leftarrow B_1 \wedge \cdots \wedge B_n$, where $n \geq 1$, A is an atom, and each B_i, $1 \leq i \leq n$, is either a literal or $\top$.
  - A is called **head** and $B_1 \wedge \cdots \wedge B_n$ **body** of the clause.
  - A clause of the form $A \leftarrow \top$ is called a **positive fact**.
  - *ground*$(\mathcal{P})$ denotes the set of all ground instances of the program $\mathcal{P}$.

$$\mathcal{P}_0 = \{p(X) \leftarrow q(X),\ p(X) \leftarrow r(X),\ r(a) \leftarrow \top\}$$
$$\rightsquigarrow \ \textit{ground}(\mathcal{P}_0) = \{p(a) \leftarrow q(a),\ p(a) \leftarrow r(a),\ r(a) \leftarrow \top\}$$

- We assume that each non-propositional program contains at least one constant.
- The language $\mathcal{L}$ underlying a program $\mathcal{P}$ shall contain precisely the relation, function and constant symbols occurring in $\mathcal{P}$, and no others.

Interpretations

- An **interpretation** is a mapping $\mathcal{L} \mapsto \{\top, \bot, \cup\}$ 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 $\bot$ and $I^\top \cap I^\bot = \emptyset$.
  - All atoms which occur neither in $I^\top$ nor $I^\bot$ are mapped to $\cup$.
- 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 $I \models \mathcal{P}$, iff $I(\mathcal{P}) = \top$.

$$\langle \emptyset, \emptyset \rangle \left\{ \begin{array}{l} \models_{Ł} \\ \not\models_{F} \end{array} \right\} \{p \leftarrow q\} = \mathcal{P}_1$$

Program Completion

- Let *ground*$(\mathcal{P})$ 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*$_1 \vee$ *Body*$_2 \vee \ldots$.
  2. If a ground atom A is not the head of any clause in *ground*$(\mathcal{P})$ then add $A \leftarrow \bot$, where $\bot$ denotes an unsatisfiable formula.
  3. All occurrences of $\leftarrow$ are replaced by $\leftrightarrow$.

  The resulting set is called **completion of *ground*$(\mathcal{P})$** or ***comp*$($*ground*$(\mathcal{P}))$**.

$$\mathcal{P}_0 = \{p(X) \leftarrow q(X),\ p(X) \leftarrow r(X),\ r(a) \leftarrow \top\}$$

$\rightsquigarrow$ *comp*$($*ground*$(\mathcal{P}_0)) = \{p(a) \leftrightarrow q(a) \vee r(a), r(a) \leftrightarrow \top,\ q(a) \leftrightarrow \bot\}$

The Fitting Operator under Fitting Semantics

- **Fitting's immediate consequence operator** $\Phi_{\mathsf{F},\mathcal{P}}(I) = \langle J^\top, J^\bot\rangle$, where

$$\begin{array}{lcl} J^\top & = & \{A \mid \text{there exists } A \leftarrow \textit{Body} \in \textit{ground}(\mathcal{P}) \text{ with } I(\textit{Body}) = \top\} \text{ and} \\ J^\bot & = & \{A \mid \text{for all } A \leftarrow \textit{Body} \in \textit{ground}(\mathcal{P}) \text{ we find } I(\textit{Body}) = \bot\}. \end{array}$$

$\mathcal{P}$	$\textit{comp}(\textit{ground}(\mathcal{P}))$	$\textit{lfp}(\Phi_{\mathsf{F},\mathcal{P}})$
$\mathcal{P}_1 = \{p \leftarrow q\}$	$\{p \leftrightarrow q,\ q \leftrightarrow \bot\}$	$\langle \emptyset, \{p, q\}\rangle$
$\mathcal{P}_2 = \{p \leftarrow q,\ q \leftarrow p\}$	$\{p \leftrightarrow q,\ q \leftrightarrow p\}$	$\langle \emptyset, \emptyset\rangle$

- $\Phi_{\mathsf{F},\mathcal{P}}$ is monotone on $(\mathcal{I}, \subseteq)$.
- If $\Phi_{\mathsf{F},\mathcal{P}}$ is continuous, then it admits a least fixed point denoted by $\textit{lfp}(\Phi_{\mathsf{F},\mathcal{P}})$.
- $\textit{lfp}(\Phi_{\mathsf{F},\mathcal{P}}) \models_{\mathsf{F}} \textit{comp}(\textit{ground}(\mathcal{P}))$.
- $\textit{lfp}(\Phi_{\mathsf{F},\mathcal{P}})$ is not necessarily a model for $\mathcal{P}$, e.g., $\langle \emptyset, \emptyset\rangle \not\models_{\mathsf{F}} \mathcal{P}_2$.
- The model intersection property does not hold, e.g.,
  $\langle \{p, q\}, \emptyset\rangle \models_{\mathsf{F}} \mathcal{P}_2$ and $\langle \emptyset, \{p, q\}\rangle \models_{\mathsf{F}} \mathcal{P}_2$ but $\langle \emptyset, \emptyset\rangle \not\models_{\mathsf{F}} \mathcal{P}_2$.

The Fitting Operator under Łukasiewicz Semantics

- *lfp*$(\Phi_{F,\mathcal{P}}) \models_Ł$ *comp*(*ground*$(\mathcal{P})$).
- If $I \models_Ł$ *comp*(*ground*$(\mathcal{P})$), then $I \models_Ł$ *ground*$(\mathcal{P})$.

$$\langle\emptyset,\emptyset\rangle \models_Ł \mathcal{P}_2 = \{p \leftarrow q,\ q \leftarrow p\}$$

- The model intersection property holds, i.e., $\cap\{I \mid I \models_Ł \mathcal{P}\} \models_Ł \mathcal{P}$.
  - If $\langle I^\top, I^\bot\rangle \models_Ł \mathcal{P}$, then $\langle I^\top, \emptyset\rangle \models_Ł \mathcal{P}$.
  - If $\langle I_1^\top, \emptyset\rangle \models_Ł \mathcal{P}$ and $\langle I_2^\top, \emptyset\rangle \models_Ł \mathcal{P}$, then $\langle I_1^\top \cap I_2^\top, \emptyset\rangle \models_Ł \mathcal{P}$.
- *lfp*$(\Phi_{F,\mathcal{P}})$ is not necessarily the least model of $\mathcal{P}$.

$$\mathcal{P} = \{p \leftarrow \top,\ q \leftarrow p,\ r \leftarrow q \wedge \neg s\}$$

$\rightsquigarrow$ *lfp*$(\Phi_{F,\mathcal{P}}) = \langle\{p,q,r\},\{s\}\rangle \neq \langle\{p,q\},\emptyset\rangle = \cap\{I \mid I \models_Ł \mathcal{P}\}$

The Stenning and van Lambalgen Operator

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

$$\begin{array}{rcl} J^\top & = & \{A \mid \text{there exists } A \leftarrow \textit{Body} \in \textit{ground}(\mathcal{P}) \text{ with } I(\textit{Body}) = \top\} \text{ and} \\ J^\bot & = & \{A \mid \text{there exists } A \leftarrow \textit{Body} \in \textit{ground}(\mathcal{P}) \text{ and} \\ & & \quad \text{for all } A \leftarrow \textit{Body} \in \textit{ground}(\mathcal{P}) \text{ we find } I(\textit{Body}) = \bot\}. \end{array}$$

- A negative fact is an expression of the form $A \leftarrow \bot$, where A is an atom.
- An extended program is a finite set of clauses and negative facts.

$\mathcal{P}$	$\mathit{lfp}(\Phi_{\mathsf{F},\mathcal{P}})$	$\mathit{lfp}(\Phi_{\mathsf{SvL},\mathcal{P}})$
$\mathcal{P}_1 = \{p \leftarrow q\}$	$\langle \emptyset, \{p, q\} \rangle$	$\langle \emptyset, \emptyset \rangle$
$\mathcal{P}_1' = \{p \leftarrow q,\ q \leftarrow \bot\}$	$\langle \emptyset, \{p, q\} \rangle$	$\langle \emptyset, \{p, q\} \rangle$

Weak Program Completion

- Let *ground*$(\mathcal{P})$ be an extended program. Consider the following transformation:

  1 All clauses with the same head $A \leftarrow Body_1,\ A \leftarrow Body_2, \ldots$ are replaced by $A \leftarrow Body_1 \vee Body_2 \vee \ldots$.

  3 All occurrences of $\leftarrow$ are replaced by $\leftrightarrow$.

  The resulting set is called weak completion of *ground*$(\mathcal{P})$ or *wcomp*$(\textit{ground}(\mathcal{P}))$.

$\mathcal{P}$	$\textit{wcomp}(\textit{ground}(\mathcal{P}))$	$\textit{lfp}(\Phi_{\mathsf{SvL},\mathcal{P}})$
$\mathcal{P}_1 = \{p \leftarrow q\}$	$\{p \leftrightarrow q\}$	$\langle \emptyset, \emptyset \rangle$
$\mathcal{P}_1' = \{p \leftarrow q,\ q \leftarrow \bot\}$	$\{p \leftrightarrow q,\ q \leftrightarrow \bot\}$	$\langle \emptyset, \{p, q\} \rangle$

The SvL Operator under Łukasiewicz Semantics

- $I_{Ł}$ is the least fixed point of $\Phi_{\text{SvL},\mathcal{P}}$ iff $I_{Ł}$ is the least model of *wcomp*(*ground*($\mathcal{P}$)).
  - Does not hold if we consider *comp*(*ground*($\mathcal{P}$)) and Fitting semantics, e.g.,

$$\mathcal{P}_1 = \{p \leftarrow q\}$$
$$\rightsquigarrow \langle \emptyset, \{p, q\}\rangle \models_{\text{F}} \textit{comp}(\mathcal{P}_1), \text{ but } \Phi_{\text{SvL},\mathcal{P}_1}(\langle \emptyset, \{p, q\}\rangle) = \langle \emptyset, \{p\}\rangle$$

  - Counterexample for Lemma 4(3) in Stenning, van Lambalgen 2008.

- *lfp*($\Phi_{\text{SvL},\mathcal{P}}$) $\models_{Ł}$ *ground*($\mathcal{P}$).
  - Does not hold if we consider Fitting semantics, e.g.,

$$\mathcal{P}_1 = \{p \leftarrow q\}$$
$$\rightsquigarrow \textit{lfp}(\Phi_{\text{Svl},\mathcal{P}_1}) = \langle \emptyset, \emptyset\rangle, \text{ but } \langle \emptyset, \emptyset\rangle \not\models_{\text{F}} \mathcal{P}_1$$

  - Counterexample for Lemma 4(1) in Stenning, van Lambalgen 2008.

The Core Method

- Various semantics for logic programs coincide with fixed points of associated immediate consequence operators (e.g., Apt, van Emden 1982).
- Banach Contraction Mapping Theorem A contraction mapping f defined on a complete metric space (X, d) has a unique fixed point. The sequence $y, f(y), f(f(y)), \ldots$ converges to this fixed point for any $y \in X$.
  - Fitting 1994 Consider logic programs, whose immediate consequence operator is a contraction.
- Funahashi 1989 Every continuous function on the reals can be uniformly approximated by feed-forward connectionist networks.
  - H., Kalinke 1994; H., Kalinke, Störr 1999 Consider logic programs, whose immediate consequence operator is continuous on the reals.

Instantiations of The Core Method

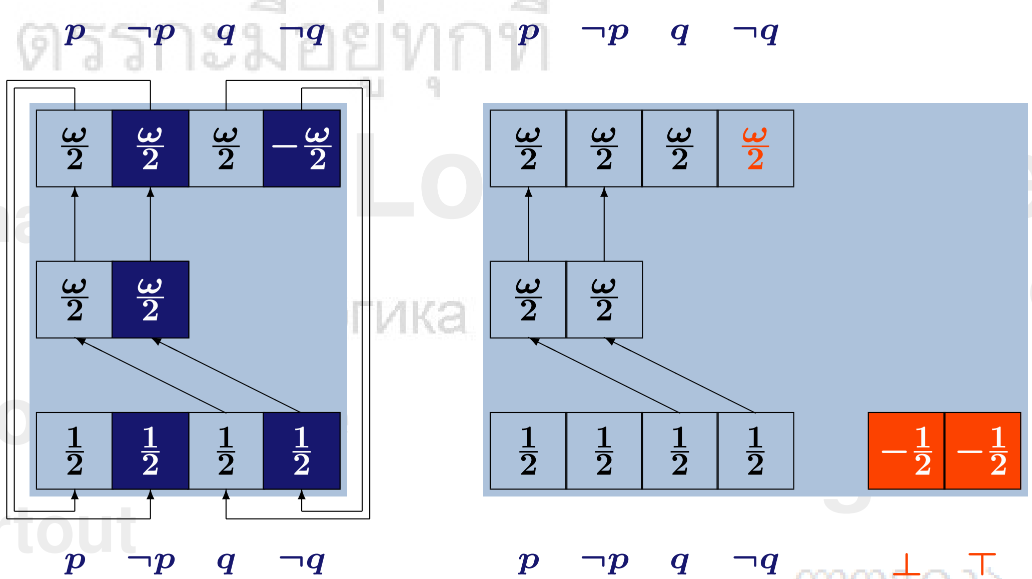
- **H., Kalinke 1994** propositional core method using binary threshold units
  - **d'Avila Garcez, Zaverucha, Carbalho 1997** bipolar sigmoidal units
  - **Kalinke 1994; Seda, Lane 2004; Kommendantskaya, Lane, Seda 2007** many-valued logic programs
  - **d'Avila Garcez, Lamb, Gabbay 2002** modal logic programs
  - **d'Avila Garcez, Lamb, Gabbay 2003** intuitionistic logic programs
  - **d'Avila Garcez, Lamb, Gabbay 2002** answer set programming, metalevel priorities
  - **d'Avila Garcez 2007** temporal reasoning
- **H., Kalinke, Störr 1999** first order logic programs
  - **Hitzler, Seda 2003; Hitzler, H., Seda 2004** many-valued logic, larger classes, other approximation theorems
  - **Bader, Hitzler, H., Witzel 2007** constructive approaches

The Core-Method for Three-Valued Logic Programs

- Consider again $\mathcal{P}_1 = \{p \leftarrow q\}$.
- A translation algorithm translates programs into a core.
- Recurrent connections connect the output to the input layer.

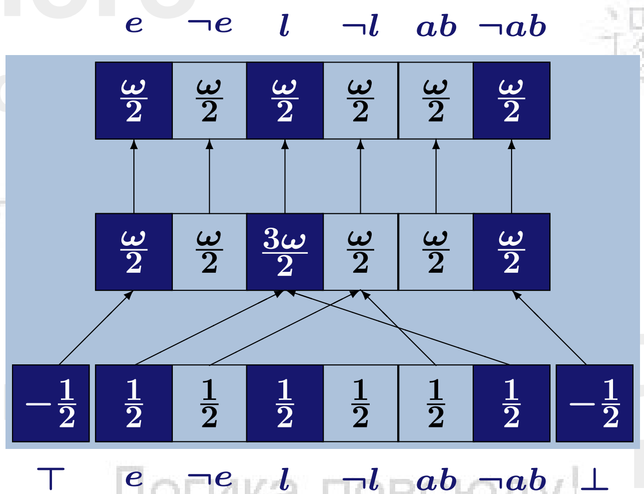
Kalinke 1995
Seda, Lane 2004 — $\Phi_{\mathsf{F},\mathcal{P}_1}$

new — $\Phi_{\mathsf{SvL},\mathcal{P}_1}$



Human Reasoning – Modus Ponens

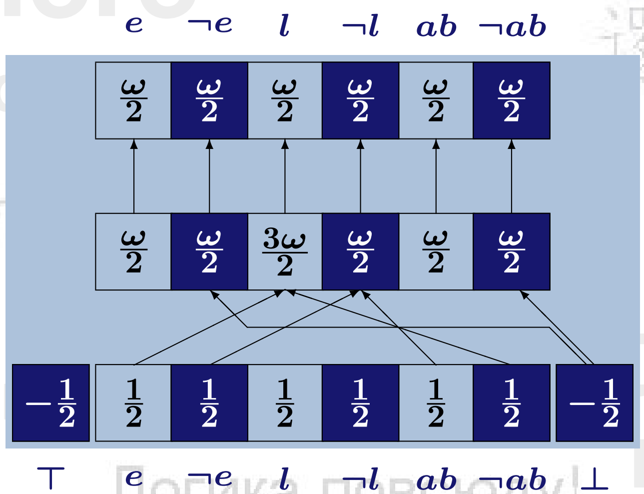
- *If Marian has an essay to write, she will study late in the library.*
  She has an essay to write.
- Byrne 1989 **96% of subjects conclude that Marian will study late in the library.**
- Stenning, van Lambalgen 2005 $\mathcal{P}_3 = \{l \leftarrow e \wedge \neg ab,\ e \leftarrow \top,\ ab \leftarrow \bot\}$.



- *lfp*$(\Phi_{\mathsf{SvL},\mathcal{P}_3}) = \langle\{l, e\}, \{ab\}\rangle$.
- **From** $\langle\{l, e\}, \{ab\}\rangle$ **follows that Marian will study late in the library.**

Human Reasoning – Denial of Antecedent (DA)

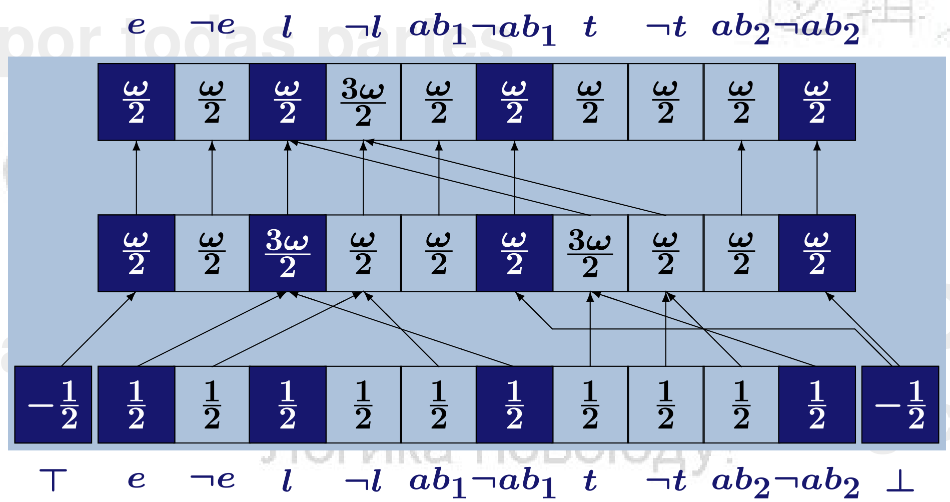
- ***If Marian has an essay to write, she will study late in the library.***
  She does not have an essay to write.
- **Byrne 1989** **46% of subjects conclude that Marian will not study late in the library.**
- **Stenning, van Lambalgen 2005** $\mathcal{P}_4 = \{l \leftarrow e \wedge \neg ab,\ e \leftarrow \bot,\ ab \leftarrow \bot\}$.



- *lfp*$(\Phi_{\mathsf{SvL},\mathcal{P}_4}) = \langle \emptyset, \{ab, e, l\}\rangle$.
- **From** $\langle \emptyset, \{ab, e, l\}\rangle$ **follows that Marian will not study late in the library.**

Human Reasoning – Alternative Arguments

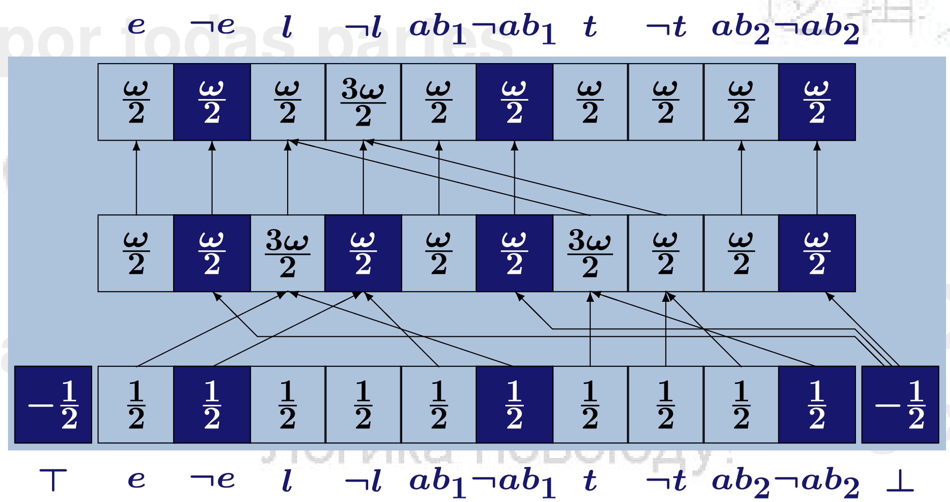
- *If Marian has an essay to write, she will study late in the library. She has an essay to write. If she has some textbooks to read, she will study late in the library.*
- Byrne 1989 96% of subjects conclude that Marian will study late in the library.
- Stenning, van Lambalgen 2005
  $\mathcal{P}_5 = \{l \leftarrow e \wedge \neg ab_1,\ e \leftarrow \top,\ ab_1 \leftarrow \bot,\ l \leftarrow t \wedge \neg ab_2,\ ab_2 \leftarrow \bot\}$.



- *lfp*$(\Phi_{\mathrm{SvL},\mathcal{P}_5}) = \langle\{e, l\}, \{ab_1, ab_2\}\rangle$.
- From $\langle\{e, l\}, \{ab_1, ab_2\}\rangle$ follows that Marian will study late in the library.

Human Reasoning – Alternative Argument and DA

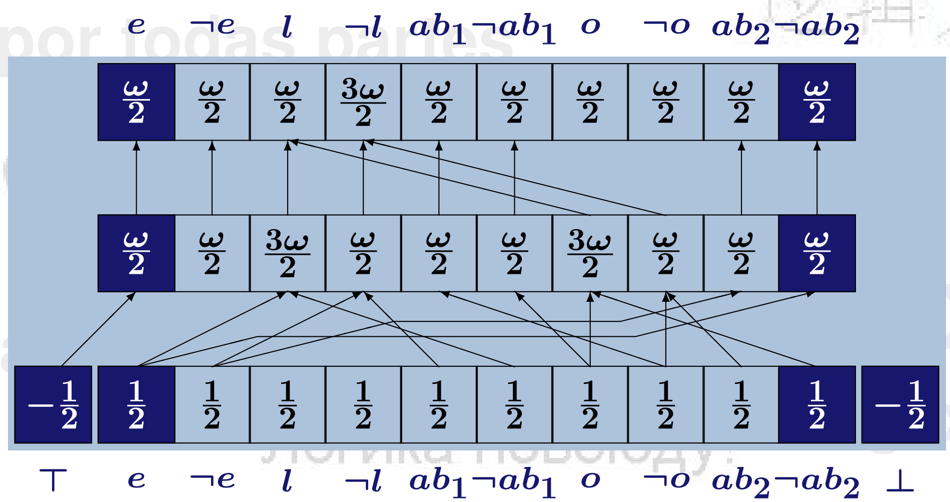
- *If Marian has an essay to write, she will study late in the library. She does not have an essay to write. If she has textbooks to read, she will study late in the library.*
- **Byrne 1989** 4% of subjects conclude that Marian will not study late in the library.
- **Stenning, van Lambalgen 2005**
  $\mathcal{P}_6 = \{l \leftarrow e \wedge \neg ab_1,\ e \leftarrow \bot,\ ab_1 \leftarrow \bot,\ l \leftarrow t \wedge \neg ab_2,\ ab_2 \leftarrow \bot\}$.



- $lfp(\Phi_{\text{SvL},\mathcal{P}_6}) = \langle \emptyset, \{ab_1, ab_2, e\}\rangle$, whereas $lfp(\Phi_{\text{F},\mathcal{P}_6}) = \langle \emptyset, \{ab_1, ab_2, e, t, l\}\rangle$.
- From $\langle \emptyset, \{ab_1, ab_2, e\}\rangle$ follows that it is unknown whether Marian will study late.

Human Reasoning – Additional Argument

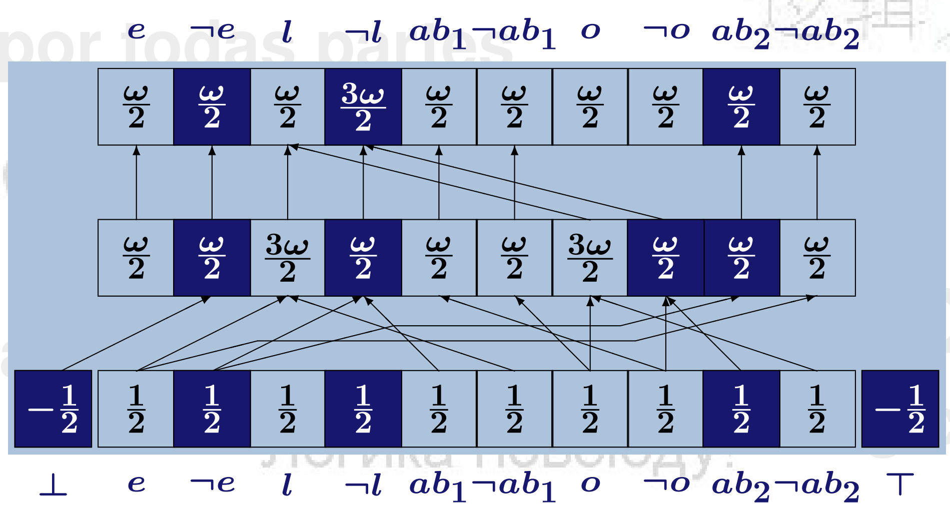
- *If Marian has an essay to write, she will study late in the library. She has an essay to write. If the library stays open, she will study late in the library.*
- **Byrne 1989** **38% of subjects conclude that Marian will study late in the library.**
- **Stenning, van Lambalgen 2005**
  $\mathcal{P}_7 = \{l \leftarrow e \wedge \neg ab_1,\ e \leftarrow \top,\ l \leftarrow o \wedge \neg ab_2,\ ab_1 \leftarrow \neg o,\ ab_2 \leftarrow \neg e, \}$.



- *lfp*$(\Phi_{\mathsf{SvL},\mathcal{P}_7}) = \langle\{e\}, \{ab_2\}\rangle$.
- **From** $\langle\{e\}, \{ab_2\}\rangle$ **follows that it is unknown whether Marian will study late.**

Human Reasoning – Additional Argument and DA

- *If Marian 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 Marian will not study late in the library.**
- Stenning, van Lambalgen 2005
  $\mathcal{P}_8 = \{l \leftarrow e \wedge \neg ab_1,\ e \leftarrow \bot,\ l \leftarrow o \wedge \neg ab_2,\ ab_1 \leftarrow \neg o,\ ab_2 \leftarrow \neg e\}.$



- *lfp*$(\Phi_{\text{SvL},\mathcal{P}_8}) = \langle\{ab_2\}, \{e, l\}\rangle.$
- **From** $\langle\{ab_2\}, \{e, l\}\rangle$ **follows that Marian will not study late in the library.**

Summary

- Three-Valued Logic Programs

Three-Valued Logic Programs	Fitting	Łukasiewicz
Model Intersection	No	Yes
Fixed Points of $\Phi_{F,\mathcal{P}}$ are models of *comp*(*ground*($\mathcal{P}$))	Yes	Yes
Fixed Points of $\Phi_{F,\mathcal{P}}$ are models of $\mathcal{P}$	No	Yes
lfp($\Phi_{SvL,\mathcal{P}}$) is the least model of *wcomp*(*ground*($\mathcal{P}$))	Yes	Yes
lfp($\Phi_{SvL,\mathcal{P}}$) is a model of $\mathcal{P}$	No	Yes

- $\Phi_{SvL,\mathcal{P}}$ can be implemented in the core method.
- Stable states of the networks correspond to least models under Łukasiewicz semantics.
- Reasoning is performed with respect to the least models under Łukasiewicz semantics and matches data from studies in human reasoning.

Discussion

- **Stenning, van Lambalgen 2005 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.**
- **Stenning, van Lambalgen 2005 also introduce integrity constraints.**
  - **If Marian has an essay to write, she will study late in the library. She will study late in the library.**
  - $\mathcal{P}_9 = \{l \leftarrow e \wedge \neg ab,\ ab \leftarrow \bot\}$.
  - **Explaining l using abduction yields e.**

Some Recent Literature

- Hölldobler, Kencana Ramli: Logic Programs under Three-Valued Łukasiwicz Semantics. ICLP 2009 (*to appear*).
- Hölldobler, Kencana Ramli: Logics and Networks for Human Reasoning. ICANN 2009 (*To appear*).
- Stenning, van Lambalgen: Human Reasoning and Cognitive Science. MIT Press 2008.