



**TECHNISCHE
UNIVERSITÄT
DRESDEN**

Faculty of Computer Science Institute of Artificial Intelligence Knowledge Representation and Reasoning

A Computational Logic Approach to the Suppression Task

Emmanuelle-Anna Dietz Steffen Hölldobler Marco Ragni

KRR Report 12-01

Mail to
Technische Universität Dresden
01062 Dresden

Bulk mail to
Technische Universität Dresden
Helmholtzstr. 10
01069 Dresden

Office
Technische Universität Dresden Room 2006
Nöthnitzer Straße 46
01187 Dresden

Internet
<http://www.wv.inf.tu-dresden.de>



**DRESDEN
concept**
Exzellenz aus
Wissenschaft
und Kultur

A Computational Logic Approach to the Suppression Task

Emmanuelle-Anna Dietz and Steffen Hölldobler (`{dietz,sh}@iccl.tu-dresden.de`)

International Center for Computation Logic, TU Dresden
D-01062 Dresden, Germany

Marco Ragni (`ragni@cognition.uni-freiburg.de`)

Center for Cognitive Science, Friedrichstraße 50
D-79098 Freiburg, Germany

Abstract

A novel approach to human conditional reasoning based on the three-valued Łukasiewicz logic is presented. We will demonstrate that the Łukasiewicz logic overcomes problems the so far proposed Fitting logic has in reasoning with the suppression task. The approach can be implemented by an appropriate connectionist network. While adequately solving the suppression task, the approach gives rise to a number of open questions concerning the use of Łukasiewicz logic, contractions, completion versus weak completion, explanations, negation, and sceptical versus credulous approaches in human reasoning.

Keywords: Łukasiewicz logic; computational logic; suppression task; human reasoning.

Introduction

Within Cognitive Science human reasoning is often studied within well-defined experiments. One of the most analyzed experiments is the *suppression task*, in which Byrne (1989) has shown that graduate students with no previous exposure to formal logic did suppress previously drawn conclusion when additional information became available. Interestingly, in some instances the previously drawn conclusions were valid whereas in other instances the conclusions were invalid with respect to classical two-valued logic. Consider the following example: *If she has an essay to write then she will study late in the library* and *If she has a textbook to read then she will study late in the library* and *She has an essay to write*. Then most participants (96%) conclude: *She will study late in the library*. If participants, however, receive, instead of the second conditional: *If the library stays open she will study late in the library* then only 38% participants conclude: *She will study late in the library*. This shows, that, although the conclusion is still correct, the conclusion is suppressed by an additional conditional. This is an excellent example for human capability to draw *non-monotonic* inferences.

Table 1 shows the abbreviations that will be used throughout the paper, whereas Table 2 gives an account of the findings of Byrne (1989).

It is straightforward to see that classical two-valued logic cannot model the suppression task adequately: Applying the classical logical consequence operator to some instances of the suppression task yields qualitatively wrong answers. At least, a non-monotonic operator is needed. As argued by Stenning and Lambalgen (2008) human reasoning should be modeled by, firstly, reasoning towards an appropriate representation and, secondly, by reasoning with respect to this representation. As appropriate representation Stenning and van

Table 1: The suppression task (Byrne, 1989) and used abbreviations. Participants received conditionals (A, B, C) and facts E, \bar{E}, L, \bar{L} and they had to draw inferences.

A	<i>If she has an essay to write then she will study late in the library.</i>
B	<i>If she has a textbook to read then she will study late in the library.</i>
C	<i>If the library stays open she will study late in the library.</i>
E	<i>She has an essay to write.</i>
\bar{E}	<i>She does not have an essay to write.</i>
L	<i>She will study late in the library.</i>
\bar{L}	<i>She will not study late in the library.</i>

Table 2: The drawn conclusions in the study of Byrne (1989).

Conditional(s)	Fact	Experimental Findings
A	E	96% of subjects conclude L.
A, B	E	96% of subjects conclude L.
A, C	E	38% of subjects conclude L.
A	\bar{E}	46% of subjects conclude \bar{L} .
A, B	\bar{E}	4% of subjects conclude \bar{L} .
A, C	\bar{E}	63% of subjects conclude \bar{L} .
A	L	53% of subjects conclude E.
A, B	L	16% of subjects conclude E.
A, C	L	55% of subjects conclude E.
A	\bar{L}	69% of subjects conclude \bar{E} .
A, B	\bar{L}	69% of subjects conclude \bar{E} .
A, C	\bar{L}	44% of subjects conclude \bar{E} .

Lambalgen propose logic programs under completion semantic based on the three-valued logic used by Fitting (1985), which itself is based on the three-valued Kleene (1952) logic.

Unfortunately, some technical claims made by Stenning and Lambalgen (2008) are wrong. It turned out, that the three-valued logic proposed by Fitting is inadequate for the suppression task. Somewhat surprisingly, the suppression task can be adequately modeled, if the three-valued Łukasiewicz (1920) logic is used. The paper gives an account of this finding and discusses a variety of open questions.

Adequacy

Computational approaches must be classified regarding cognitive adequacy. In particular, we can distinguish between conceptual and inferential measures. In our context, a system is *conceptually adequate* if it appropriately represents human knowledge. *Inferential adequacy* measures whether the computations behave similarly to human reasoning. In cognitive science theories are evaluated by performing reasoning experiments on subjects. For instance, Knauff (1999) investigate which kind of information humans use when representing and remembering spatial arrangements in Allen’s interval calculus. In computer science, one commonly used hypothesis is, that if computational models are biologically plausible then they should also behave similar to the biological brain (Herrmann & Ohl, 2009). However, until now there are no implemented models which easily process computations given a large amounts of data or efficiently deal with incomplete information. These aspects are fundamental for elementary reasoning processes. Shastri and Ajjanagadde (1993) present a connectionist approach for reflexive reasoning called SHRUTI and state that their system is psychologically plausible. Furthermore, Beringer and Hölldobler (1993) conclude from a logical reconstruction of SHRUTI that “*adequacy implies massive parallelism.*” In this paper, we evaluate the adequacy of our computational logic approach by examining that our approach qualitatively gives the same answers as subjects in the suppression task experiments.

A Computational Logic Approach

Stenning and Lambalgen (2008) have proposed to use logic programs under completion semantics and based on a three-valued logic to model the suppression task. In particular, they suggest that human reasoning is modeled by, firstly, reasoning towards an appropriate representation or logical form (conceptual adequacy) and, secondly, reasoning with respect to this representation (inferential adequacy).

In the following we introduce three-valued logics and, in particular, the Łukasiewicz logic. As the chosen representation are logic programs, such programs are introduced next together with their (weak) completion. We adopt the reasoning step towards an appropriate logical form from Stenning and Lambalgen (2008). Thereafter, we discuss three-valued models for logic programs under the Łukasiewicz semantics and, in particular, the model intersection property which entails the existence of least models. We show that the conclusions drawn with respect to these least models correspond to the findings in (Byrne, 1989) and conclude that the derived logics programs under Łukasiewicz semantics are conceptually adequate for the suppression task.

In order to investigate inferential adequacy we consider the semantic operator associated with logic programs as defined by Stenning and Lambalgen (2008). For each program \mathcal{P} , this operator admits a least fixed point, which is equal to the least Łukasiewicz model of \mathcal{P} . At this point we are able discuss the technical problems in (Stenning & Lambalgen, 2008), while

showing that they do not occur if we use Łukasiewicz semantics. The least fixed point of such a semantic operator can be computed within a connectionist setting, where the application of the operator to some interpretation requires only two steps in time and the time to compute the least model is linear in the number of reasoning steps an agent has to perform.

Finally, we add abduction to the approach in order and show that sceptical reasoning is needed in order to model the suppression task adequately.

Three-Valued Logics

Three-valued logics were introduced by Łukasiewicz (1920). In Table 3 the truth tables of his logic are depicted, where \top , \perp , and U denote *true*, *false*, and *unknown*, respectively.

Table 3: The three-valued Łukasiewicz logic.

	\neg		\wedge	\vee	\leftarrow	\leftrightarrow
\top	\perp	\top	\top	\top	\top	\top
\perp	\top	\top	\perp	\top	\top	\perp
U	U	\top	U	\top	\top	U
		\perp	\perp	\top	\perp	\perp
		\perp	\perp	\perp	\top	\top
		\perp	U	U	U	U
		U	\top	\top	U	U
		U	\perp	U	\top	U
		U	U	U	\top	\top

With the introduction of the third truth value, there are plenty of options to define the truth tables for the connectives. For example, Kleene (1952) introduced an implication, whose truth table is identical to the Łukasiewicz implication except that the case where precondition and conclusion are both mapped to \perp : in this case, the implication itself is mapped to U . Kleene also introduced a so-called *strong equivalence*, where the truth value \top is assigned to $F \leftrightarrow G$ if F and G are assigned to identical truth values, and \perp is assigned otherwise. Fitting (1985) combined the truth tables for \neg , \vee , \wedge from Łukasiewicz with the Kleene implication and strong equivalence for investigations within logic programming. We will call this combination the *Fitting semantics*.¹ Stenning and Lambalgen (2008) use Fitting semantics without giving a reason for this particular choice.

Logic Programs

A (*logic*) *program* is a finite set of expressions of the form

$$A \leftarrow B_1 \wedge \dots \wedge B_n, \quad (1)$$

where $n \geq 1$, A is an atom, and each B_i , $1 \leq i \leq n$, is either a literal, \top , or \perp . A is called *head* and $B_1 \wedge \dots \wedge B_n$ is called *body* of the *clause* (1). A clause of the form $A \leftarrow \top$ is called

¹We believe that Fitting had termination analysis of logic programs in his mind when he selected this particular logic.

positive fact, whereas a clause of the form $A \leftarrow \perp$ is called negative fact. In the sequel, let \mathcal{P} be a program.

Consider the following transformation for a given \mathcal{P} :

1. All clauses with the same head $A \leftarrow Body_1, A \leftarrow Body_2, \dots$ are replaced by $A \leftarrow Body_1 \vee Body_2 \vee \dots$
2. If an atom A is not the head of any clause in \mathcal{P} (and, thus, is *undefined* in \mathcal{P}) then add $A \leftarrow \perp$.
3. All occurrences of \leftarrow are replaced by \leftrightarrow .

The resulting set is called *completion* of \mathcal{P} ($c\mathcal{P}$). If step 2 is omitted, then the resulting set is called *weak completion* of \mathcal{P} ($wc\mathcal{P}$). It is well-known that reasoning with respect to the completion of a logic program is non-monotonic.

Reasoning Towards an Appropriate Logical Form

Stenning and Lambalgen (2008) have argued that the first step in modeling human reasoning is reasoning towards an appropriate logical form. In particular, they argue that conditionals shall not be encoded by implications straight away but rather by licenses for implications. For example, the conditional A should be encoded by the clause $l \leftarrow e \wedge \overline{ab_1}$, where ab_1 is an *abnormality* predicate which expresses that something abnormal is known. In other words, l holds if e holds and nothing abnormal is known.

In this paper, we simply adopt this reasoning step from Stenning and Lambalgen (2008). In the first two columns of Table 4 the programs obtained for the first six examples of the suppression task are depicted. The third column shows the weak completions of the programs.

Three-Valued Models for Logic Programs

A (*three-valued*) *interpretation* is a mapping from a propositional language to the set $\{\top, \perp, \text{U}\}$ of truth values. It is quite common to represent interpretations by tuples of the form $\langle I^\top, I^\perp \rangle$, where I^\top contains all atoms which are mapped to \top , I^\perp contains all atoms which are mapped to \perp , $I^\top \cap I^\perp = \emptyset$, and all atoms which occur neither in I^\top nor in I^\perp are mapped to U .

Let \mathcal{P} be a program and I an interpretation. I is a (*three-valued*) *model under Łukasiewicz semantics* for \mathcal{P} ($I \models_{3L} \mathcal{P}$) iff each clause occurring in \mathcal{P} is mapped to \top using the truth table depicted in Table 3. Likewise, \models_{3F} can be defined with respect to the Fitting semantics.

In Hölldobler and Kencana Ramli (2009b) it was shown that the model intersection property holds for (weakly completed) programs under Łukasiewicz semantics, i.e.,

$$\begin{aligned} \cap\{I \mid I \models_{3L} \mathcal{P}\} &\models_{3L} \mathcal{P}, \\ \cap\{I \mid I \models_{3L} wc\mathcal{P}\} &\models_{3L} wc\mathcal{P}. \end{aligned}$$

The model intersection property for programs does not hold under Fitting semantics: Let $\mathcal{P} = \{p \leftarrow q\}$, then both, $\langle \{p, q\}, \emptyset \rangle$ and $\langle \emptyset, \{p, q\} \rangle$, are models for \mathcal{P} , whereas $\langle \emptyset, \emptyset \rangle$ is not a model for \mathcal{P} .

Table 4: A summary of the computational logic approach to the suppression task (Part 1).

\mathcal{P}	clauses	$wc\mathcal{P}$	$\text{lm } wc\mathcal{P}$
\mathcal{P}_{AE}	$l \leftarrow e \wedge \overline{ab_1}$ $ab_1 \leftarrow \perp$ $e \leftarrow \top$	$l \leftrightarrow e \wedge \overline{ab_1}$ $ab_1 \leftrightarrow \perp$ $e \leftrightarrow \top$	$\langle \{e, l\}, \{ab_1\} \rangle$
\mathcal{P}_{ABE}	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow t \wedge \overline{ab_2}$ $ab_1 \leftarrow \perp$ $ab_2 \leftarrow \perp$ $e \leftarrow \top$	$l \leftrightarrow (e \wedge \overline{ab_1}) \vee (t \wedge \overline{ab_2})$ $ab_1 \leftrightarrow \perp$ $ab_2 \leftrightarrow \perp$ $e \leftrightarrow \top$	$\langle \{e, l\}, \{ab_1, ab_2\} \rangle$
\mathcal{P}_{ACE}	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow o \wedge \overline{ab_3}$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$ $e \leftarrow \top$	$l \leftrightarrow (e \wedge \overline{ab_1}) \vee (o \wedge \overline{ab_3})$ $ab_1 \leftrightarrow \overline{o}$ $ab_3 \leftrightarrow \overline{e}$ $e \leftrightarrow \top$	$\langle \{e\}, \{ab_3\} \rangle$
$\mathcal{P}_{A\overline{E}}$	$l \leftarrow e \wedge \overline{ab_1}$ $ab_1 \leftarrow \perp$ $e \leftarrow \perp$	$l \leftrightarrow e \wedge \overline{ab_1}$ $ab_1 \leftrightarrow \perp$ $e \leftrightarrow \perp$	$\langle \emptyset, \{e, l, ab_1\} \rangle$
$\mathcal{P}_{A\overline{B}\overline{E}}$	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow t \wedge \overline{ab_2}$ $ab_1 \leftarrow \perp$ $ab_2 \leftarrow \perp$ $e \leftarrow \perp$	$l \leftrightarrow (e \wedge \overline{ab_1}) \vee (t \wedge \overline{ab_2})$ $ab_1 \leftrightarrow \perp$ $ab_2 \leftrightarrow \perp$ $e \leftrightarrow \perp$	$\langle \emptyset, \{e, ab_1, ab_2\} \rangle$
$\mathcal{P}_{A\overline{C}\overline{E}}$	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow o \wedge \overline{ab_3}$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$ $e \leftarrow \perp$	$l \leftrightarrow (e \wedge \overline{ab_1}) \vee (o \wedge \overline{ab_3})$ $ab_1 \leftrightarrow \overline{o}$ $ab_3 \leftrightarrow \overline{e}$ $e \leftrightarrow \perp$	$\langle \{ab_3\}, \{e, l\} \rangle$

The model intersection property guarantees the existence of least models for logic programs as well as for their weak completions. Column 4 in Table 4 depicts the least models for the weak completions of the programs encoding the first six examples of the selection task, where lm denotes the least model of its argument (under Łukasiewicz semantics).

Reasoning with Respect to Least Models

Because programs as well as their weak completions admit the model intersection property we can reason wrt the least models. Returning to the first six examples of the suppression task we find

$$\begin{aligned} \text{lm}_{3LWC} \mathcal{P}_{AE} &= \langle \{e, l\}, \{ab_1\} \rangle && \models_{3L} l \\ \text{lm}_{3LWC} \mathcal{P}_{ABE} &= \langle \{e, l\}, \{ab_1, ab_2\} \rangle && \models_{3L} l \\ \text{lm}_{3LWC} \mathcal{P}_{ACE} &= \langle \{e\}, \{ab_3\} \rangle && \not\models_{3L} l \vee \overline{l} \\ \text{lm}_{3LWC} \mathcal{P}_{A\overline{E}} &= \langle \emptyset, \{e, l, ab_1\} \rangle && \models_{3L} \overline{l} \\ \text{lm}_{3LWC} \mathcal{P}_{A\overline{B}\overline{E}} &= \langle \emptyset, \{e, ab_1, ab_2\} \rangle && \not\models_{3L} l \vee \overline{l} \\ \text{lm}_{3LWC} \mathcal{P}_{A\overline{C}\overline{E}} &= \langle \{ab_3\}, \{e, l\} \rangle && \models_{3L} \overline{l} \end{aligned}$$

Compared to the experimental findings presented in Table 2, the presented approach appears to be adequate from a qualitative point of view.

Computing Least Models

In Computational Logic, least models are usually computed as least fixed points of appropriate semantic operators (see, e.g., Apt & Emden, 1982). Stenning and Lambalgen (2008) devised such an operator for the class of programs discussed herein: Let I be an interpretation in $\Phi_{\mathcal{P}}(I) = \langle J^{\top}, J^{\perp} \rangle$, where

$$\begin{aligned} J^{\top} &= \{A \mid \text{there exists } A \leftarrow \text{body} \in \mathcal{P} \text{ with } I(\text{body}) = \text{true}\}, \\ J^{\perp} &= \{A \mid \text{there exists } A \leftarrow \text{body} \in \mathcal{P} \text{ and} \\ &\quad \text{for all } A \leftarrow \text{body} \in \mathcal{P} \text{ we find } I(\text{body}) = \text{false}\}. \end{aligned}$$

As shown in Hölldobler and Kencana Ramli (2009b) for any \mathcal{P} , the least fixed point of $\Phi_{\mathcal{P}}$ is identical to $\text{lm}_{3LWC} \mathcal{P}$ and can be computed by iterating $\Phi_{\mathcal{P}}$ starting with the empty interpretation. Moreover, as shown in Hölldobler and Kencana Ramli (2009c) the least fixed point of $\Phi_{\mathcal{P}}$ can be computed by a recurrent neural network with a feed-forward core.

One should observe that in this paper $\Phi_{\mathcal{P}}$ uses Łukasiewicz semantics whereas in Stenning and Lambalgen (2008) it uses Fitting semantics. The difference is striking if we consider the least fixed point of $\Phi_{\mathcal{P}_{ACE}}$, which is $\langle \{e\}, \{ab_3\} \rangle$ under both semantics. Whereas under Łukasiewicz semantics this fixed point is a model for \mathcal{P}_{ACE} , under Fitting semantics the clause $l \leftarrow o \wedge \bar{ab}_3 \in \mathcal{P}_{ACE}$ is mapped to \perp . This is a counter example for Lemma 4(1.) in Stenning and Lambalgen (2008).

Now consider the case that we use Fitting semantics and the completion of $\mathcal{P}_{AB\bar{E}}$. The least fixed point of $\Phi_{\mathcal{P}_{AB\bar{E}}}$ is $\langle \emptyset, \{e, ab_1, ab_2\} \rangle$. Note that $\langle \emptyset, \{e, ab_1, ab_2\} \rangle \not\models_{3F} \mathcal{P}_{AB\bar{E}}$ because under completion t must be mapped to \perp and, hence, l will be mapped to \perp as well. This is a counter example for Lemma 4(3.) in Stenning and Lambalgen (2008). The example also shows that reasoning under the Fitting semantics and wrt the completion of a program is not adequate as only 4% of humans conclude \bar{l} in this case.

Contractions

As mentioned in the previous subsection, the least fixed point of the operator $\Phi_{\mathcal{P}}$ can be computed by iterating $\Phi_{\mathcal{P}}$ starting with the empty interpretation. However, if the operator is a contraction, then by the Banach Contraction Theorem (Banach, 1922) the operator has a unique fixed point which can be computed by iterating the operator starting with an arbitrary interpretation. As shown in Hölldobler and Kencana Ramli (2009a), $\Phi_{\mathcal{P}}$ is a contraction if \mathcal{P} is acyclic, i.e., if there is a mapping l from the set of atoms to \mathbb{N} such that for each clause $A \leftarrow \text{Body} \in \mathcal{P}$ and each $B \in \text{Body}$ we find $l(A) > l(B)$. One should observe that all programs shown in Table 4 are acyclic using, for example, the following mapping:

atom	\perp	\top	t	o	e	ab_3	ab_2	ab_1	l
$l(\text{atom})$	1	1	2	3	4	5	6	7	8

Abduction

The second part of the suppression task deals with the affirmation of the consequent and modus tollens. These reasoning

processes can best be described as abductive, that is, a plausible explanation is computed given some observation. Following Kakas, Kowalski, and Toni (1993) we consider an *abductive framework* consisting of a program \mathcal{P} as knowledge base, a set \mathcal{A} of abducibles consisting of the (positive and negative) facts for each undefined predicate symbol in \mathcal{P} ,² and the logical consequence relation \models_{3L}^{lmwc} , where $\mathcal{P} \models_{3L}^{lmwc} F$ iff $\text{lm}_{3LWC} \mathcal{P}(F) = \top$. As *observations* we consider literals.

Let $\langle \mathcal{P}, \mathcal{A}, \models_{3L}^{lmwc} \rangle$ be an abductive framework and O an observation. O is *explained* by \mathcal{E} iff $\mathcal{E} \subseteq \mathcal{A}$, $\mathcal{P} \cup \mathcal{E}$ is satisfiable, and $\mathcal{P} \cup \mathcal{E} \models_{3L}^{lmwc} O$. Usually, minimal explanations are preferred. In case there exist several minimal explanations, then two forms of reasoning can be distinguished. F follows *sceptically* from program \mathcal{P} and observation O ($\mathcal{P}, O \models_s F$) iff O can be explained and for all minimal explanation \mathcal{E} we find $\mathcal{P} \cup \mathcal{E} \models_{3L}^{lmwc} O$, whereas F follows *credulously* from \mathcal{P} and O ($\mathcal{P}, O \models_c F$) iff there exists a minimal explanation \mathcal{E} such that $\mathcal{P} \cup \mathcal{E} \models_{3L}^{lmwc} O$.³

Table 5 depicts the programs, the observations and the minimal explanations for the second part of the suppression task in the second, third, and fourth row, respectively. The final row shows the least model of the weak completion of the union of the program and the minimal explanation under the Łukasiewicz semantics. If we reason sceptically wrt to these least models, then we obtain

$$\begin{array}{ll} \mathcal{P}_A, l \models_s e, & \mathcal{P}_A, \bar{l} \models_s \bar{e}, \\ \mathcal{P}_{AB}, l \not\models_s e, & \mathcal{P}_{AB}, \bar{l} \models_s \bar{e}, \\ \mathcal{P}_{AC}, l \models_s e, & \mathcal{P}_{AC}, \bar{l} \not\models_s \bar{e}, \end{array}$$

which are qualitatively adequate answers if compared to Table 2. One should observe that a credulous agent concludes e from $\mathcal{P} = \mathcal{P}_{AB}$ and $O = l$, which according to Byrne (1989) only 16% of the tested subjects did.

Open Questions

Łukasiewicz Logic

This logic was selected because the technical bugs in Stenning and Lambalgen (2008) can be solved by switching from Fitting to Łukasiewicz semantics. In particular, the model intersection property holds under Łukasiewicz semantics. Hence, for each program \mathcal{P} a least model does exist which can be computed as least fixed point of the associated semantic operator $\Phi_{\mathcal{P}}$. Moreover, a rigorous study has revealed that the suppression task can be adequately modeled under Łukasiewicz semantics, whereas this does not hold for Fitting semantics. Nevertheless, the main question of whether Łukasiewicz logic is adequate for human reasoning is still open. For example, in the Łukasiewicz logic the semantic deduction theorem does not hold. Hence, it would be interesting to see how humans deal with the deduction theorem. Can other typical human reasoning problems like the Wason (1968) selection task be adequately modeled under Łukasiewicz semantics?

²Recall that A is *undefined* in \mathcal{P} iff \mathcal{P} does not contain a clause of the form $A \leftarrow \text{Body}$.

³See (Hölldobler, Philipp, & Wernhard, 2011) for more details.

Table 5: A summary of the computational logic approach to the suppression task (Part 2). The cases $\mathcal{P} = \mathcal{P}_{AB}, O = l$ and $\mathcal{P} = \mathcal{P}_{AC}, O = \bar{l}$ have two minimal extensions.

\mathcal{P}	clauses	O	\mathcal{E}	$\text{lm}_{3L\text{wc}}(\mathcal{P} \cup \mathcal{E})$
\mathcal{P}_A	$l \leftarrow e \wedge \overline{ab_1}$ $ab_1 \leftarrow \perp$	l	$e \leftarrow \top$	$\langle \{e, l\}, \{ab_1\} \rangle$
\mathcal{P}_{AB}	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow t \wedge \overline{ab_2}$ $ab_1 \leftarrow \perp$ $ab_2 \leftarrow \perp$	l	$e \leftarrow \top$	$\langle \{e, l\}, \{ab_1, ab_2\} \rangle$
			$t \leftarrow \top$	$\langle \{l, t\}, \{ab_1, ab_2\} \rangle$
\mathcal{P}_{AC}	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow o \wedge \overline{ab_3}$ $ab_1 \leftarrow \bar{o}$ $ab_3 \leftarrow \bar{e}$	l	$e \leftarrow \top$ $o \leftarrow \top$	$\langle \{e, l, o\}, \{ab_1, ab_3\} \rangle$
\mathcal{P}_A	$l \leftarrow e \wedge \overline{ab_1}$ $ab_1 \leftarrow \perp$	\bar{l}	$e \leftarrow \perp$	$\langle \emptyset, \{e, l, ab_1\} \rangle$
\mathcal{P}_{AB}	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow t \wedge \overline{ab_2}$ $ab_1 \leftarrow \perp$ $ab_2 \leftarrow \perp$	\bar{l}	$e \leftarrow \perp$ $t \leftarrow \perp$	$\langle \emptyset, \{e, l, t, ab_1, ab_2\} \rangle$
\mathcal{P}_{AC}	$l \leftarrow e \wedge \overline{ab_1}$ $l \leftarrow o \wedge \overline{ab_3}$ $ab_1 \leftarrow \bar{o}$ $ab_3 \leftarrow \bar{e}$	\bar{l}	$e \leftarrow \perp$	$\langle \{ab_3\}, \{e, l\} \rangle$
			$o \leftarrow \perp$	$\langle \{ab_1\}, \{o, l\} \rangle$

Contractions

For each program \mathcal{P} shown in Table 4 the operator $\Phi_{\mathcal{P}}$ is a contraction. Thus, there is a unique fixed point, which can be computed by iterating $\Phi_{\mathcal{P}}$ on some initial interpretation. Consequently, if in the suppression task subjects are influenced towards some initial non-empty interpretation, their performance should not differ provided that they have enough time to compute the least fixed point; it should differ, however, if they are interrupted before the least fixed point is computed and asked to reason wrt the interpretation computed so far. Another aspect is about level mapping. It might have the additional function to represent some ordering about the subjects knowledge. For instance, it is easy to see that $l(t), l(o)$, and $l(e)$ have to be smaller than $l(ab_1), l(ab_2)$, and $l(ab_3)$ to enforce acyclicity. As l does not occur in the body for any clause, $l(l)$ is mapped to the highest level. For human reasoning that means l does not imply any further knowledge. On the other hand, we have $l(\top)$ and $l(\perp)$ which are never the head of any clause and therefore are mapped to the smallest level. They both have the function to express facts.

Completion versus Weak Completion

The program $\mathcal{P}_{AB\bar{E}}$ served as an example to illustrate that completion is inadequate for the suppression task whereas weak completion is adequate. Likewise, Hölldobler et al.

(2011) have shown in a detailed study that the programs mentioned in Table 5 together with their minimal explanations must be weakly completed in order to adequately model the suppression task, whereas completion does not. Are there other human reasoning episodes which support the claim that weak completion is adequate? Even if so, the problem remains to explicitly add negative facts (in the reasoning step towards an appropriate logical form) for those predicates, which should be mapped to \perp like ab_1 in the program \mathcal{P}_{AE} .

Sceptical versus Credulous Reasoning

The case of program $\mathcal{P} = \mathcal{P}_{AB}$ and observation $O = l$ in Table 5 shows that agents must reason sceptically in order to adequately model this case. Whereas this is a striking case for sceptical reasoning, the case $\mathcal{P} = \mathcal{P}_{AC}$ and $O = \bar{l}$ is less convincing. A sceptical agent will not conclude \bar{e} , whereas a credulous agent will conclude \bar{e} . Compared to the corresponding case (A, C, \bar{L}) shown in Table 2, 44% of the subjects conclude \bar{E} . Unfortunately, Byrne (1989) (and related publications that we are aware of) give no account of the distribution of the answers given by those subjects who did not conclude \bar{E} . Hence, at the moment we can argue in favor of a sceptical agent (*the majority of the subjects did not conclude \bar{E}*), but – given the complete distribution – it may be the case that one can argue in favor of a credulous agent (*there are more subjects concluding \bar{E} than subjects concluding E and subjects answering “I don’t know”*).

In this context, it might be useful to explicitly differentiate between *inferential* knowledge and facts. For a credulous agent the amount of *inferential* knowledge does not influence its conclusion. On the other hand, for a sceptical agent, as more *inferential* knowledge is given, as more supporting facts are necessary to draw some conclusion.

Explanations

The approach presented in this paper is based on minimal explanations. Although, there are findings corroborating the human preference of minimal explanations (over non-minimal ones) (Ormerod, Manktelow, & Jones, 1993) – this holds only partially (Johnson-Laird, Girotto, & Legrenzi, 2004). Computational models of abduction typically generate explanations iteratively such that minimal explanations are generated first. How are minimal explanations computed by humans? What happens, if there are more than one minimal explanation?

Negation

In the presented approach positive information is preferred over negative one. Consider, for example, the program $\mathcal{P} = \{q \leftarrow \top, q \leftarrow \perp\}$. The least model of $\text{wc}\mathcal{P}$ is $\langle \{q\}, \emptyset \rangle$ and, hence, an agent reasoning wrt to this model will conclude q . Is this consistent with human reasoning? The presented approach could be extended to include integrity constraints like $\perp \leftarrow q$. Any model for a program containing such an integrity constraint must map q to \perp . Is this adequate for human reasoning? If so, under which conditions shall such integrity

constraints be added within the reasoning step towards an appropriate logical form?

Connectionist Realization

As shown in (Hölldobler & Kencana Ramli, 2009c), the computation of the least fixed point of the semantic operator $\Phi_{\mathcal{P}}$ associated with a program \mathcal{P} can be realized within the core-method (Bader, Hitzler, Hölldobler, & Witzel, 2007). In this connectionist realization, $\Phi_{\mathcal{P}}$ is computed by a feed-forward network, whose output units are recurrently connected to the input units. Whereas this network is trainable by backpropagation and, thus, $\Phi_{\mathcal{P}}$ can be learned by experience, there is no evidence whatsoever that backpropagation is biological plausible. The approach can be extended to handle abduction following (Garcez, Gabbay, Ray, & Woods, 2007). However, in this setting, explanations are generated in a fixed, hard-wired sequence, which does not seem to be plausible either.

Summary

We have presented an adequate computational logic approach for the suppression task. It is based on weakly completed logic programs under Łukasiewicz semantics. Such programs admit least models which can be computed by iterating an appropriate semantic operator. Reasoning is performed wrt the least models. The approach is extended by sceptical reasoning within an abductive framework. Moreover, it can be realized in a connectionist setting. The approach has been carefully tested against alternatives like completed logic programs, Fitting semantics, and credulous reasoning, but none of these variations was found to be adequate.

Acknowledgments

Many thanks to Bertram Fronhöfer, Caroline Dewi Puspa Kencana Ramli, Tobias Philipp, and Christoph Wernhard.

References

- Apt, K., & Emden, M. van. (1982). Contributions to the theory of logic programming. *J. of the ACM*, 29, 841-862.
- Bader, S., Hitzler, P., Hölldobler, S., & Witzel, A. (2007). The core method: Connectionist model generation for first-order logic programs. In B. Hammer & P. Hitzler (Eds.), *Perspectives of neural-symbolic integration* (Vol. 77, p. 205-232). Berlin, Heidelberg: Springer.
- Banach, S. (1922). Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fund. Math.*, 3, 133-181.
- Beringer, A., & Hölldobler, S. (1993). On the adequateness of the connection method. In *Proceedings of the aaai national conference on artificial intelligence* (p. 9-14).
- Byrne, R. (1989). Suppressing valid inferences with conditionals. *Cognition*, 31, 61-83.
- Fitting, M. (1985). A Kripke-Kleene semantics for logic programs. *Journal of Logic Programming*, 2(4), 295-312.
- Garcez, A. d'Avila, Gabbay, D., Ray, O., & Woods, J. (2007). Abductive reasoning in neural-symbolic learning systems. *TOPOI*, 26, 37-49.
- Herrmann, C. S., & Ohl, F. W. (2009). Cognitive adequacy in brain-like intelligence. In B. Sendhoff, E. Körner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-Like Intelligence*. Springer.
- Hölldobler, S., & Kencana Ramli, C. D. P. (2009a). Contraction properties of a semantic operator for human reasoning. In L. Li & K. K. Yen (Eds.), *Proceedings of the fifth international conference on information* (p. 228-231). International Information Institute.
- Hölldobler, S., & Kencana Ramli, C. D. P. (2009b). Logic programs under three-valued Łukasiewicz's semantics. In P. M. Hill & D. S. Warren (Eds.), *Logic programming* (Vol. 5649, p. 464-478). Springer Berlin Heidelberg.
- Hölldobler, S., & Kencana Ramli, C. D. P. (2009c). Logics and networks for human reasoning. In C. Alippi, M. M. Polycarpou, C. G. Panayiotou, & G. Ellinasetal (Eds.), *Artificial neural networks – ICANN* (Vol. 5769, p. 85-94). Springer Berlin Heidelberg.
- Hölldobler, S., Philipp, T., & Wernhard, C. (2011). An abductive model for human reasoning. In *Proceeding tenth international symposium on logical formalizations of commonsense reasoning*. (commonsensereasoning.org/2011/proceedings.html)
- Johnson-Laird, P., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, 111(3), 640-661.
- Kakas, A. C., Kowalski, R. A., & Toni, F. (1993). Abductive Logic Programming. *Journal of Logic and Computation*, 2(6), 719-770.
- Kleene, S. (1952). *Introduction to metamathematics*. North-Holland.
- Knauff, M. (1999). The cognitive adequacy of Allen's interval calculus for qualitative spatial representation and reasoning. *Spatial Cognition and Computation*, 1, 261-290.
- Łukasiewicz, J. (1920). O logice trójwartościowej. *Ruch Filozoficzny*, 5, 169-171. (English translation: On Three-Valued Logic. In: *Jan Łukasiewicz Selected Works*. (L. Borkowski, ed.), North Holland, 87-88, 1990.)
- Ormerod, T., Manktelow, K., & Jones, G. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology*, 46A(4), 653-677.
- Shastri, L., & Ajjanagadde, V. (1993). From associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioural and Brain Sciences*, 16(3), 417-494.
- Stenning, K., & Lambalgen, M. van. (2008). *Human reasoning and cognitive science*. MIT Press.
- Wason, P. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, 20, 273-281.