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. While adequately solving the suppression task, the approach gives rise to a number of open questions concerning the use of Łukasiewicz logic, unique fixed points, completion versus weak completion, explanations, negation, and sceptical versus credulous approaches in human reasoning.

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

Introduction

An interesting study is the *suppression task*, in which Byrne (1989) has shown that graduate students with no previous exposure to formal logic did suppress previously drawn conclusions 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 finish then she will study late in the library* and *She has an essay to finish.* Then most subjects (96%) conclude: *She will study late in the library.* If subjects, however, receive an additional conditional: *If the library stays open she will study late in the library* then only 38% of the subjects 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). As we are using a formal language, propositions like "She will go to the library" (abbreviated L) will be represented by propositional variables like l, with the intended interpretation that if l is *true* ($\top$), then "She will go to the library". Taking a naive propositional approach, we can represent A by the implication $e \leftarrow l$, where the propositional variables e and l represent the facts E and L, respectively, and so on.

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 (like A, C, E) yields qualitatively wrong answers, due to the monotonic nature of the classical logic.

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

A	*If she has an essay to finish then she will study late in the library.*
B	*If she has a textbook to read then she will study late in the library.*
C	*If the library stays open she will study late in the library.*
E	*She has an essay to finish.*
$\overline{E}$	*She does not have an essay to finish.*
L	*She will study late in the library.*
$\overline{L}$	*She will not study late in the library.*

Table 2: The drawn conclusions in the experiment of Byrne.

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	$\overline{E}$	46% of subjects conclude $\overline{L}$.
A, B	$\overline{E}$	4% of subjects conclude $\overline{L}$.
A, C	$\overline{E}$	63% of subjects conclude $\overline{L}$.
A	L	53% of subjects conclude E.
A, B	L	16% of subjects conclude E.
A, C	L	55% of subjects conclude E.
A	$\overline{L}$	69% of subjects conclude $\overline{E}$.
A, B	$\overline{L}$	69% of subjects conclude $\overline{E}$.
A, C	$\overline{L}$	44% of subjects conclude $\overline{E}$.

Consequently, at least a non-monotonic operator is needed. As argued by Stenning and van Lambalgen (2008)[1] human reasoning should be modeled by, first, reasoning towards an appropriate representation and, second, by reasoning with respect to this representation. As appropriate representation Stenning and van Lambalgen propose logic programs under completion semantics 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 van Lambalgen (2008) are wrong concerning their second step. Hölldobler and Kencana Ramli (2009b) have shown 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

[1]There is an earlier publication (Stenning & van Lambalgen, 2005), but Michiel van Lambalgen advised us to refer to their textbook.

Łukasiewicz (1920) logic is used. The paper gives an account of this finding and discusses a variety of consequences of this new logic and some open questions.

Adequacy

Computational approaches to explain human reasoning should be *cognitively adequate*. Usually, the concept of adequacy is measured by distinguishing between conceptual and inferential adequacy (Strube, 1996). 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. It is common in Cognitive Science to evaluate theories by performing reasoning experiments on subjects. For instance, Knauff (1999) investigates 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. In this paper, we evaluate the inferential adequacy of our computational logic approach by examining that our approach gives the same answers as subjects in the suppression task experiments.

A Computational Logic Approach

Stenning and van 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, first, reasoning towards an appropriate representation or logical form (conceptual adequacy) and, second, 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 van 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 logic programs under Łukasiewicz semantics are inferentially 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 van 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 to discuss the technical problems in (Stenning & van Lambalgen, 2008), while showing that they do not occur if we use Łukasiewicz semantics. Finally, we add abduction to the approach 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). Table 3 gives the truth tables for different three-valued logics. The symbols $\top$, $\bot$, and U denote *true*, *false*, and *unknown*, respectively. For instance, if F is mapped to $\bot$ and G is mapped to $\top$ then their conjunction ($F \wedge G$) is mapped to $\bot$ and their disjunction ($F \vee G$) is mapped to $\top$. By introducing

Table 3: The three-valued logics

F	$\overline{F}$
$\top$	$\bot$
$\bot$	$\top$
U	U

F	G	$\wedge$	$\vee$	$\leftarrow_L$	$\leftrightarrow_L$	$\leftarrow_K$	$\leftrightarrow_S$
$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$	$\top$	$\top$	$\bot$	$\top$	$\bot$
$\top$	U	U	$\top$	$\top$	U	$\top$	$\bot$
$\bot$	$\top$	$\bot$	$\top$	$\bot$	$\bot$	$\bot$	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$	$\top$	$\top$	$\top$	$\top$
$\bot$	U	$\bot$	U	U	U	U	$\bot$
U	$\top$	U	$\top$	U	U	U	$\bot$
U	$\bot$	$\bot$	U	$\top$	U	$\top$	$\bot$
U	U	U	U	$\top$	$\top$	U	$\top$

a third truth value, there are various options to define the truth tables for the connectives. For example, Kleene (1952) introduced an implication ($\leftarrow_K$), whose truth table is identical to the Łukasiewicz implication ($\leftarrow_L$) except in the cases where precondition and conclusion are both mapped to U: In this case, the implication itself is mapped to U by Kleene, but to $\top$ by Łukasiewicz. The set of connectives under Łukasiewicz semantics is $\{\neg, \wedge, \vee, \leftarrow_L, \leftrightarrow_L\}$. Kleene also introduced a so-called *strong equivalence*, where the truth value $\top$ is assigned to $F \leftrightarrow_S G$ if F and G are assigned to identical truth values, and $\bot$ 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* where the set of connectives is $\{\neg, \wedge, \vee, \leftarrow_K, \leftrightarrow_S\}$[2]. Stenning and van 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

$$A \leftarrow B_1 \wedge \ldots \wedge B_n, \tag{1}$$

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 $B_1 \wedge \ldots \wedge B_n$ is called *body* of the *clause* (1). A clause of the form $A \leftarrow \top$ is called *positive fact*, whereas a clause of the form $A \leftarrow \bot$ is called *negative fact*. In the sequel, $\mathcal{P}$ shall denote a logic program.

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

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

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

The resulting set is called *completion* of $\mathcal{P}$ ($\mathsf{c}\,\mathcal{P}$). If step 2 is omitted, then the resulting set is called *weak completion* of $\mathcal{P}$ ($\mathsf{wc}\,\mathcal{P}$). Consider $\mathcal{P} = \{p \leftarrow q\}$, then $\mathsf{c}\,\mathcal{P} = \{p \leftrightarrow q, q \leftrightarrow \bot\}$. $\mathsf{c}\,\mathcal{P}$ entails that p and q are mapped to $\bot$. Reasoning with respect to the completion of a logic program is non-monotonic. For instance, if $\mathcal{P}' = \mathcal{P} \cup \{q \leftarrow \top\}$, then $\mathsf{c}\,\mathcal{P}'$ entails that p and q are mapped to $\top$. The process of weak completion can be associated with the human interpretation of conditionals as biconditionals (Evans, Newstead & Byrne, 1993).

Reasoning Towards an Appropriate Logical Form

Stenning and van 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.

We think that Stenning and van Lambalgen (2008) adequately model the representational part of the suppression task and adopt this reasoning step. Our focus is on the inferential aspect of their approach. In the first two columns of Table 4 the programs obtained for the first six examples of the suppression task are depicted. For instance, in $\mathcal{P}_{ACE}$ we have that *She will study late in the library* if either *She has an essay to finish* and *Nothing abnormal (ab_1) is known* or *She has a textbook to read* and *Nothing abnormal (ab_3) is known*. The predicates ab_1, ab_2 and ab_3 represent different kinds of abnormality. For instance, ab_1 is true when *The library does not stay open* and ab_3 is true when *She does not have an essay to finish*.

Three-Valued Models for Logic Programs

A (*three-valued*) *interpretation* is a mapping from a propositional language to the set $\{\top, \bot, \mathsf{U}\}$ of truth values. It is quite common to represent interpretations by tuples of the form $\langle I^\top, I^\bot \rangle$, where $I^\top$ contains all atoms which are mapped to $\top$, $I^\bot$ contains all atoms which are mapped to $\bot$, $I^\top \cap I^\bot = \emptyset$, and all atoms which occur neither in $I^\top$ nor in $I^\bot$ are mapped to U. Let $\mathcal{P}$ be a program and I an interpretation. I is a (*three-valued*) *model under Łukasiewicz semantics for* $\mathcal{P}$ ($I \models_L \mathcal{P}$) if and only if each clause occurring in $\mathcal{P}$ is mapped to $\top$ using the truth table depicted in Table 3. Likewise, $\models_F$ can be defined with respect to the Fitting semantics. For instance, consider $\mathcal{P} = \{p \leftarrow q\}$. Then under Łukasiewicz semantics we have three different models $\langle\{p,q\},\emptyset\rangle$, $\langle\emptyset,\{p,q\}\rangle$ and $\langle\emptyset,\emptyset\rangle$. Under Fitting semantics only the first two interpretations are models, because if p and q are mapped to U then $p \leftarrow q \in \mathcal{P}$ is mapped to U as well.

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

$\mathcal{P}$	clauses	$\mathsf{wc}\,\mathcal{P}$	$\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}$	*Byrne*
$\mathcal{P}_{AE}$	$l \leftarrow e \wedge \overline{ab}_1$ $ab_1 \leftarrow \bot$ $e \leftarrow \top$	$l \leftrightarrow e \wedge \overline{ab}_1$ $ab_1 \leftrightarrow \bot$ $e \leftrightarrow \top$	$\langle\{e,l\},\{ab_1\}\rangle$	96% L
$\mathcal{P}_{ABE}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow t \wedge \overline{ab}_2$ $ab_1 \leftarrow \bot$ $ab_2 \leftarrow \bot$ $e \leftarrow \top$	$l \leftrightarrow (e \wedge \overline{ab}_1)$ $\vee (t \wedge \overline{ab}_2)$ $ab_1 \leftrightarrow \bot$ $ab_2 \leftrightarrow \bot$ $e \leftrightarrow \top$	$\langle\{e,l\},\{ab_1,ab_2\}\rangle$	96% L
$\mathcal{P}_{ACE}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow o \wedge \overline{ab}_3$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$ $e \leftarrow \top$	$l \leftrightarrow (e \wedge \overline{ab}_1)$ $\vee (o \wedge \overline{ab}_3)$ $ab_1 \leftrightarrow \overline{o}$ $ab_3 \leftrightarrow \overline{e}$ $e \leftrightarrow \top$	$\langle\{e\},\{ab_3\}\rangle$	38% L
$\mathcal{P}_{A\overline{E}}$	$l \leftarrow e \wedge \overline{ab}_1$ $ab_1 \leftarrow \bot$ $e \leftarrow \bot$	$l \leftrightarrow e \wedge \overline{ab}_1$ $ab_1 \leftrightarrow \bot$ $e \leftrightarrow \bot$	$\langle\emptyset,\{e,l,ab_1\}\rangle$	46% $\overline{L}$
$\mathcal{P}_{AB\overline{E}}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow t \wedge \overline{ab}_2$ $ab_1 \leftarrow \bot$ $ab_2 \leftarrow \bot$ $e \leftarrow \bot$	$l \leftrightarrow (e \wedge \overline{ab}_1)$ $\vee (t \wedge \overline{ab}_2)$ $ab_1 \leftrightarrow \bot$ $ab_2 \leftrightarrow \bot$ $e \leftrightarrow \bot$	$\langle\emptyset,\{e,ab_1,ab_2\}\rangle$	4% $\overline{L}$
$\mathcal{P}_{AC\overline{E}}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow o \wedge \overline{ab}_3$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$ $e \leftarrow \bot$	$l \leftrightarrow (e \wedge \overline{ab}_1)$ $\vee (o \wedge \overline{ab}_3)$ $ab_1 \leftrightarrow \overline{o}$ $ab_3 \leftrightarrow \overline{e}$ $e \leftrightarrow \bot$	$\langle\{ab_3\},\{e,l\}\rangle$	63% $\overline{L}$

Reasoning with Respect to Least Models

In order to identify the desired model of a certain program, we reason with respect to their least models. Least models are guaranteed to exist if the model intersection property holds:

$$\cap\{I \mid I \models_L \mathcal{P}\} \models_L \mathcal{P},$$
$$\cap\{I \mid I \models_L \mathsf{wc}\,\mathcal{P}\} \models_L \mathsf{wc}\,\mathcal{P}.$$

In Hölldobler and Kencana Ramli (2009b) it was shown that the model intersection property holds for (weakly completed) programs under Łukasiewicz semantics. The model intersection property for programs does not hold under Fitting semantics. Consider again $\mathcal{P} = \{p \leftarrow q\}$, then both, $\langle\{p,q\},\emptyset\rangle$ and $\langle\emptyset,\{p,q\}\rangle$, are models for $\mathcal{P}$, whereas their intersection $\langle\emptyset,\emptyset\rangle$ is not a model for $\mathcal{P}$ under Fitting semantics.

The third column of Table 4 shows the weak completions of the programs encoding the first six examples of the suppression task. Column 4 in Table 4 depicts the corresponding least models where lm_L denotes the least model of its argument under Łukasiewicz semantics. The last column shows the results of the suppression task. Specifically we find that

$$\begin{array}{lcll}
\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}_{AE} & = & \langle\{e,l\},\{ab_1\}\rangle & \models_L \; l \\
\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}_{ABE} & = & \langle\{e,l\},\{ab_1,ab_2\}\rangle & \models_L \; l \\
\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}_{ACE} & = & \langle\{e\},\{ab_3\}\rangle & \not\models_L \; l \vee \overline{l} \\
\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}_{A\overline{E}} & = & \langle\emptyset,\{e,l,ab_1\}\rangle & \models_L \; \overline{l} \\
\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}_{AB\overline{E}} & = & \langle\emptyset,\{e,ab_1,ab_2\}\rangle & \not\models_L \; l \vee \overline{l} \\
\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}_{AC\overline{E}} & = & \langle\{ab_3\},\{e,l\}\rangle & \models_L \; \overline{l}
\end{array}$$

where $\not\models_L$ means that a given formula cannot be concluded. Our approach coincides with the seemingly favored results of the suppression task and thus appears to be adequate.

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 van Lambalgen (2008) devised such an operator for programs discussed herein: Let I be an interpretation in $\Phi_{\mathcal{P}}(I) = \langle J^\top, J^\bot \rangle$, where

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

As shown in Hölldobler and Kencana Ramli (2009b) for any $\mathcal{P}$, the least fixed point of $\Phi_{\mathcal{P}}$ is identical to $\mathsf{lm}_{3L}\mathsf{wc}\,\mathcal{P}$ and can be computed by iterating $\Phi_{\mathcal{P}}$ starting with the empty interpretation. The following example shows how the least model of $\mathcal{P}_{ACE}$ is computed starting with interpretation $I_0 = \langle \emptyset, \emptyset \rangle$:

$$\begin{array}{lclcllcl} I_1 & = & \Phi_{\mathcal{P}_{ACE}}(I_0) & = & \langle \{e\}, \emptyset \rangle & & & \\ I_2 & = & \Phi_{\mathcal{P}_{ACE}}(I_1) & = & \langle \{e\}, \{ab_3\} \rangle & = & \Phi_{\mathcal{P}_{ACE}}(I_2) \end{array}$$

where I_2 is the least fixed point of $\Phi_{\mathcal{P}_{ACE}}$. This is not a model under Fitting semantics because the clause $l \leftarrow o \wedge ab_3 \in \mathcal{P}_{ACE}$ is mapped to U and not to $\top$ such as under Łukasiewicz semantics. This is a counter example for Lemma 4(1.) in Stenning and van Lambalgen (2008) which states that the least fixed point of the $\Phi_{\mathcal{P}}$ operator under Fitting semantics is the minimal model of $\mathcal{P}$. Another statement made by Stenning and van Lambalgen (2008), Lemma 4(3.) states, that all models of $\mathsf{c}\,\mathcal{P}$ are fixed points of $\Phi_{\mathcal{P}}$ and every fixed point is a model. Consider the completion of $\Phi_{\mathcal{P}_{AB\overline{E}}}$, then t is mapped to $\bot$ and therefore l is mapped to $\bot$ as well. However, its least fixed point is $\langle \emptyset, \{e, ab_1, ab_2\} \rangle$ where t and l are undefined. This example also shows that reasoning under Fitting semantics and with respect to the completion of a program is not adequate as only 4% conclude $\overline{L}$ in this case.

Unique Fixed Point

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[3]. All programs shown in Table 4 are acyclic.

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 $\mathcal{P}$,[4] and the logical consequence relation $\models_L^{lmwc}$, where $\mathcal{P} \models_L^{lmwc} F$ if and only if $\mathsf{lm}_{3L}\mathsf{wc}\,\mathcal{P}(F) = \top$ for the formula F. As *observations* we consider literals.

Let $\langle \mathcal{P}, \mathcal{A}, \models_L^{lmwc} \rangle$ be an abductive framework and O an observation. O is *explained* by $\mathcal{E}$ if and only if $\mathcal{E} \subseteq \mathcal{A}$, $\mathcal{P} \cup \mathcal{E}$ is satisfiable, and $\mathcal{P} \cup \mathcal{E} \models_L^{lmwc} O$. 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 ($\mathcal{P}, O \models_s F$) if and only if O can be explained and for all minimal explanations $\mathcal{E}$ we find $\mathcal{P} \cup \mathcal{E} \models_L^{lmwc} O$, whereas F follows *credulously* from $\mathcal{P}$ and O ($\mathcal{P}, O \models_c F$) if and only if there exists a minimal explanation $\mathcal{E}$ such that $\mathcal{P} \cup \mathcal{E} \models_L^{lmwc} O$.[5] For instance, consider the following two programs under sceptical reasoning:

1. $\mathcal{P}_{AB}$ where $O = l$: $\mathcal{A} = \{e \leftarrow \top, e \leftarrow \bot, t \leftarrow \top, t \leftarrow \bot\}$ and $\mathsf{lm}_{3L}\mathsf{wc}\,\mathcal{P}_{AB} = \langle \emptyset, \{ab_1, ab_2\} \rangle$. There are two minimal explanations with either $\{e \leftarrow \top\}$ and $\{t \leftarrow \top\}$. Thus, we cannot conclude whether *She has an essay to finish* or not.
2. $\mathcal{P}_{AC}$ where $O = l$: $\mathcal{A} = \{e \leftarrow \top, e \leftarrow \bot, o \leftarrow \top, o \leftarrow \bot\}$ and $\mathsf{lm}_{3L}\mathsf{wc}\,\mathcal{P}_{AC} = \langle \emptyset, \emptyset \rangle$. There is only one minimal explanation $\{e \leftarrow \top, o \leftarrow \top\}$ and thus *She has an essay to finish.*

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 column, respectively. The second last column shows the least model of the weak completion of the union of the program and the minimal explanation under the Łukasiewicz semantics and the final one shows the results of the suppression task. If we reason sceptically with respect to these least models, then we obtain

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

which are adequate answers if compared to the seemingly favored results of the suppression task. 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 subjects did.

Open Questions

Łukasiewicz Logic

This logic was selected because the technical bugs in Stenning and van Lambalgen (2008) can be solved by switching from Fitting to Łukasiewicz semantics. In particular, the

[3] A program $\mathcal{P}$ is acyclic if there exists a numbering for all propositional variables such that for all clauses in $\mathcal{P}$ the value of the head is strictly larger than the value of the literals in the body.

[4] Recall that A is *undefined* in $\mathcal{P}$ if and only if $\mathcal{P}$ does not contain a clause of the form $A \leftarrow Body$.

[5] See (Hölldobler, Philipp, & Wernhard, 2011) for more details.

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

$\mathcal{P}$	clauses	$\mathcal{O}$	$\mathcal{E}$	$\mathsf{lm}_L\mathsf{wc}\,(\mathcal{P} \cup \mathcal{E})$	*Byrne*
$\mathcal{P}_A$	$l \leftarrow e \wedge \overline{ab}_1$ $ab_1 \leftarrow \bot$	l	$e \leftarrow \top$	$\langle\{e, l\}, \{ab_1\}\rangle$	53% E
$\mathcal{P}_{AB}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow t \wedge \overline{ab}_2$ $ab_1 \leftarrow \bot$ $ab_2 \leftarrow \bot$	l	$e \leftarrow \top$	$\langle\{e, l\}, \{ab_1, ab_2\}\rangle$	16% E
			$t \leftarrow \top$	$\langle\{l, t\}, \{ab_1, ab_2\}\rangle$	
$\mathcal{P}_{AC}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow o \wedge \overline{ab}_3$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$	l	$e \leftarrow \top$ $o \leftarrow \top$	$\langle\{e, l, o\}, \{ab_1, ab_3\}\rangle$	55% E
$\mathcal{P}_A$	$l \leftarrow e \wedge \overline{ab}_1$ $ab_1 \leftarrow \bot$	$\bar{l}$	$e \leftarrow \bot$	$\langle\emptyset, \{e, l, ab_1\}\rangle$	69% $\overline{E}$
$\mathcal{P}_{AB}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow t \wedge \overline{ab}_2$ $ab_1 \leftarrow \bot$ $ab_2 \leftarrow \bot$	$\bar{l}$	$e \leftarrow \bot$ $t \leftarrow \bot$	$\langle\emptyset, \{e, l, t, ab_1, ab_2\}\rangle$	69% $\overline{E}$
$\mathcal{P}_{AC}$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow o \wedge \overline{ab}_3$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$	$\bar{l}$	$e \leftarrow \bot$	$\langle\{ab_3\}, \{e, l\}\rangle$	44% $\overline{E}$
			$o \leftarrow \bot$	$\langle\{ab_1\}, \{l, o\}\rangle$	

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 Deduction Theorem does not hold [6]. 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?

Unique Fixed Point

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 with respect to the interpretation computed so far.

Completion versus Weak Completion

The program $\mathcal{P}_{AB\overline{E}}$ 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 $\bot$ like ab_1 in the program $\mathcal{P}_{AE}$.

Sceptical versus Credulous Reasoning

The case of program $\mathcal{P} = \mathcal{P}_{AB}$ and observation $\mathcal{O} = l$ 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 $\mathcal{O} = \bar{l}$ is less convincing. A sceptical agent will not conclude $\bar{e}$, whereas a credulous agent will conclude $\bar{e}$. Compared to the corresponding case $(A, C, \overline{L})$ shown in Table 2, 44% of the subjects conclude $\overline{E}$. Unfortunately, Byrne (1989) (and related publications that we are aware of) gives no account of the distribution of the answers given by those subjects who did not conclude $\overline{E}$. Hence, at the moment we can argue in favor of a sceptical agent (*the majority of the subjects did not conclude* $\overline{E}$), 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* $\overline{E}$ *than subjects concluding E and subjects answering "I don't know"*).

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 $\mathcal{P} = \{q \leftarrow \top, q \leftarrow \bot\}$. The least model of $\mathsf{wc}\,\mathcal{P}$ is $\langle\{q\}, \emptyset\rangle$ and, hence, an agent reasoning with respect to this model will conclude q. Is this consistent with human reasoning? The presented approach could be extended to include integrity constraints like $\bot \leftarrow q$. Any model for a program containing such an integrity constraint must map q to $\bot$. Is this adequate for human reasoning? If so, under which conditions shall such

[6] A logic satisfies the *Deduction Theorem* if for any finite set of formulae $\Phi = \{\phi_1, \phi_2, \ldots, \phi_n\}$ and any formula ψ the following holds: $\Phi \models \psi$ if and only if $\models (\phi_1 \wedge \phi_2 \wedge \cdots \wedge \phi_n) \rightarrow \psi$.

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 with respect to 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.

Byrne, R. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 61-83.

Evans, J.S.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ England: Lawrence Erlbaum Associates.

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 *Proceedingth 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*.

Stenning, K., & Lambalgen, M. van. (2005). Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, *29*(6), 919-960.

Stenning, K., & Lambalgen, M. van. (2008). *Human reasoning and cognitive science*. MIT Press.

Strube, G. (1996). *Wörterbuch der Kognitionswissenschaft*. Klett-Cotta.

Wason, P. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*, 273-281.