A Computational Logic Approach to the Abstract and the Social Case of the Selection Task

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

Previous results have shown that weak completion semantics based on three-valued Łukasiewicz logic can adequately represent and explain human behavior in the suppression task. This approach corresponds to well-founded semantics for tight logic programs. In this paper we apply both semantics to the selection task – probably the most famous and best investigated psychological study about human reasoning with conditionals. In its abstract version, cards are shown to some people and they have to check if a conditional statement about the cards holds true. Numerous psychological studies show that most people do not solve this task correctly in terms of classical propositional logic and tend to make similar reasoning errors. Once the same reasoning problem is framed within a social setting, most people solve the task correctly. By distinguishing between belief and social constraints, we apply the abstract and the social case within the weak completion and the well-founded semantics and show that when reasoning towards the corresponding representations, our computational approach adequately reflects the psychological results. Finally, we present a psychological study testing different predictions of the weak completion and the well-founded semantics on programs which are not tight.

Introduction

In the last century the classical (propositional) logic calculus has played an important role as a normative concept for psychologists investigating human reasoning. Psychological research, however, showed that humans systematically deviate from the logically correct answers. Some attempts to formalize such behavior have already been made in the field of Computational Logic such as in non-monotonic logic, common sense reasoning or three-valued logics, where incomplete information is expressible. Furthermore, the fields of Artificial Neural Networks and Cognitive Science focus on challenging problems that aim to simulate and understand human reasoning.

Computational approaches that try to explain human reasoning should be evaluated according to their *cognitive adequacy*. The concept of adequacy has been defined in a linguistic context to compare and explain language theories and their properties (Strube, 1996). Two different adequacy measures are defined: *conceptual adequacy* and *inferential adequacy*. Conceptual adequacy reflects in how far a language represents a content correctly. Inferential adequacy is about the procedural part how language is applied to content (Strube, 1996). In Computational Logic, the interpretation of these measurements can be understood as follows: conceptual adequacy deals with the representational part of the system. The aim is to have a representation of the given information such that it captures the structure of how it appears in human knowledge. Inferential adequacy measures whether the computations are similar to the way humans reason. Analogously, Stenning and van Lambalgen (2008) argue that human reasoning should be modeled by, first, reasoning towards an appropriate representation and, second, by reasoning with respect to this representation.

As appropriate representation for modeling the suppression task, Stenning and van Lambalgen (2008) 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 are wrong. Hölldobler and Kencana Ramli (2009; 2009) have shown that the three-valued logic proposed by Stenning and van Lambalgen is inadequate for the suppression task. Somewhat surprisingly, the suppression task can be adequately modeled if the three-valued Łukasiewicz logic presented in (Łukasiewicz, 1920) is used instead. The computational logic approach (Hölldobler & Kencana Ramli, 2009; Dietz, Hölldobler, & Ragni, 2012) models the suppression task as logic programs under the so-called weak completion, a variation of Clark's (Clark, 1978) completion. They show that these conclusions drawn with respect to least models correspond to the findings of Byrne (1989) and conclude that the derived logic programs under Łukasiewicz logic are inferentially adequate for the suppression task. Furthermore, Dietz, Hölldobler, and Wernhard (2013) show that there is a strong correspondence between weak completion and well-founded semantics (Van Gelder, Ross, & Schlipf, 1991) for the class of tight programs.

In this paper, we apply our approach to another psychological study, the *Wason selection task* (Wason, 1968). In the Wason selection task participants had to check a given conditional statement on some instances. The problem was presented as a rather abstract description and almost all participants made the same classical logical mistakes. Griggs and Cox (1982) developed an isomorphic representation of the problem in a social context, and surprisingly almost all of the participants solved this task correctly. Kowalski (2011) gives an interesting interpretation of this difference, which we will use for our approach.

In the following we briefly review three-valued logics and give the necessary definitions for weak completion semantics. After that, we explain the Wason selection task and our computational logic approach. Finally, we present results from a psychological experiment to evaluate whether well-founded or weak completion semantics is more adequate.

Three-valued Logics

Three-valued logics were first conceived by Łukasiewicz (1920) and since then different interpretations of the connectives have been proposed. The corresponding truth values are $\top$, $\bot$ and U, which mean *true*, *false* and and *unknown*, respectively. Kleene (1952) introduced an implication ($\leftarrow_K$), whose truth table is identical to Łukasiewicz implication ($\leftarrow_Ł$) 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 mapped to $\top$ by Łukasiewicz. The set of connectives under Łukasiewicz semantics is $\{\neg, \wedge, \vee, \leftarrow_Ł, \leftrightarrow_Ł\}$.

A further common variant of three-valued implication ($\leftarrow_S$) is called seq_3 in Gottwald (2001). The corresponding equivalence ($\leftrightarrow_S$) assigns $\top$ to $F \leftrightarrow G$ if and only if F and G are mapped to identical truth values, and $\bot$ is assigned otherwise. Fitting (1985) combined the truth tables for $\neg$, $\vee$ and $\wedge$ from Łukasiewicz with the equivalence $\leftrightarrow_S$ for investigations within Logic Programming. Hence, the set of connectives used by Fitting is $\{\neg, \wedge, \vee, \leftrightarrow_S\}$. Table 1 gives the truth tables of three-valued conjunction, disjunction and the different variations of implication and equivalence.

Stenning and van Lambalgen (2008) modeled the suppression task by extending the logic used by Fitting with $\leftarrow_K$. Hölldobler and Kenana Ramli (2009) showed that this logic is inadequate and proposed to use Łukasiewicz semantics which corrects some technical mistakes and adequately models the suppression task.

Under well-founded semantics the interpretation of the implication corresponds to $\leftarrow_S$ (Przymusinski, 1989), which corresponds to the interpretation of the implication in the logic S_3 (Rescher, 1969), that is $\{\neg, \wedge, \vee, \leftarrow_S, \leftrightarrow_S\}$. As indicated by the highlighted $\top$ signs in Table 1, whenever a formula is true under $\leftarrow_S$ then it is true under $\leftarrow_L$, and vice versa. The underlying three-valued logic for weak completion semantics which we present in the following, corresponds to three-valued Łukasiewicz logic.

Preliminaries

We define the necessary notations we will use throughout this paper and restrict ourselves to propositional logic as this is sufficient for our purpose. A *logic program* $\mathcal{P}$ is a finite set of clauses of the form

$$A \leftarrow A_1 \wedge \cdots \wedge A_n \wedge \neg B_1 \wedge \cdots \wedge \neg B_m \qquad (1)$$

where A is an atom called *head* and $A_1 \wedge \cdots \wedge A_n \wedge \neg B_1 \wedge \cdots \wedge \neg B_m$ is called *body* of the clause and A_i, with $1 \leq i \leq n$, and B_j, with $1 \leq j \leq m$, are atoms. $\top$ and $\bot$ are special atoms where $A \leftarrow \top$ expresses the *fact* that A is true and $A \leftarrow \bot$ expresses the *negative fact* that A is false. [1] Without loss of generality we assume that the bodies of clauses are not empty and restrict the use of $\top$ and $\bot$ to facts as indicated. $\mathsf{Atoms}(\mathcal{P})$ denotes the set of all atoms occurring in the program $\mathcal{P}$. An atom A is *defined* in $\mathcal{P}$ if there is a clause with head A; otherwise it is said to be *undefined* in $\mathcal{P}$; $\mathsf{ud}\,(\mathcal{P}) = \{A \mid \text{there is no clause } C \text{ in } \mathcal{P} \text{ such that } A \text{ is the head of } C\}$ is the set of undefined atoms in $\mathcal{P}$. A *normal logic program* is a logic program without negative facts. If $\mathcal{P}$ is a logic program then $\mathcal{P}^+$ denotes the program obtained from $\mathcal{P}$ by deleting all negative facts.

Program Classes

The following three programs can be classified with respect to whether they contain cycles:

$$\begin{array}{ccc} \mathcal{P}_1 & \mathcal{P}_2 & \mathcal{P}_3 \\ \{p \leftarrow q\} & \{p \leftarrow q,\ q \leftarrow p\} & \{p \leftarrow \neg q,\ q \leftarrow \neg p\} \end{array}$$

Cycles occur in programs when at least one atom depends on itself: we say that p *depends on* q if and only if there exists a clause $p \leftarrow A_1 \wedge \cdots \wedge A_n \wedge \neg B_1 \wedge \cdots \wedge \neg B_m$such that $q = A_i$ or $q = B_j$ for some $1 \leq i \leq n$ or $1 \leq j \leq m$. Dependency is transitive, thus if p depends on q and q depends on r, then p depends on r. We distinguish between two types of dependency: p *depends positively* on q if $q = A_i$ and p *depends negatively* on q if $q = B_j$ where one negative dependency is sufficient to define the whole dependency as negative. We have a *cycle* in a program if at least one atom depends on itself. If the dependency is positive, then it is a *positive cycle*, otherwise it is a *negative cycle*.

Accordingly, we distinguish between the following program classes: *Acyclic* programs do not contain cycles. $\mathcal{P}_1$ is an acyclic program, whereas $\mathcal{P}_2$ and $\mathcal{P}_3$ are not. *Stratified* programs (Apt, Blair, & Walker, 1988) only contain positive cycles. $\mathcal{P}_2$ is a stratified program, but $\mathcal{P}_3$ is not. *Tight* programs (Erdem & Lifschitz, 2003) only contain negative cycles. $\mathcal{P}_3$ is a tight program, but $\mathcal{P}_2$ is not.

Interpretations and Models

An *interpretation* I is a mapping from formulas to the set of truth values $\{\top, \bot, \mathsf{U}\}$. The truth value of a given formula under a given interpretation is determined according to the corresponding three-valued logic. We represent an interpretation as a pair $I = \langle I^\top, I^\bot \rangle$ of disjoint sets of atoms where $I^\top$ is the set of all atoms that are mapped to $\top$ by I and $I^\bot$ is the set of all atoms that are mapped to $\bot$ by I. If atoms are mapped to U, they are neither in $I^\top$ nor in $I^\bot$. A *total* interpretation with respect to a program $\mathcal{P}$ is an interpretation $I = \langle I^\top, I^\bot \rangle$ such that $\mathsf{Atoms}(\mathcal{P}) = I^\top \cup I^\bot$.

One should observe that in contrast to two-valued logic, $A \leftarrow B$ and $A \vee \neg B$ are not semantically equivalent, neither for $\leftarrow_Ł$ nor for $\leftarrow_S$. For example, consider the case $I(A) = I(B) = \mathsf{U}$. Then, $I(A \vee \neg B) = \mathsf{U}$ whereas

$$I(A \leftarrow_Ł B) = I(A \leftarrow_S B) = \top.$$

[1] The notion of falsehood appears to be counterintuitive at first sight, but programs will be interpreted under completion semantics where the implication sign is replaced by an equivalence sign.

Table 1: Truth tables for three-valued logics. The highlighted $\top$'s indicate that formulas of the form $A \leftarrow B$ which are true under $\leftarrow_L$ are true under $\leftarrow_S$, and vice versa.

F	$\neg F$
$\top$	$\bot$
$\bot$	$\top$
U	U

$\wedge$	$\top$	U	$\bot$
$\top$	$\top$	U	$\bot$
U	U	U	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$

$\vee$	$\top$	U	$\bot$
$\top$	$\top$	$\top$	$\top$
U	$\top$	U	U
$\bot$	$\top$	U	$\bot$

$\leftarrow_Ł$	$\top$	U	$\bot$
$\top$	$\top$	$\top$	$\top$
U	U	$\top$	$\top$
$\bot$	$\bot$	U	$\top$

$\leftrightarrow_Ł$	$\top$	U	$\bot$
$\top$	$\top$	U	$\bot$
U	U	$\top$	U
$\bot$	$\bot$	U	$\top$

$\leftarrow_S$	$\top$	U	$\bot$
$\top$	$\top$	$\top$	$\top$
U	$\bot$	$\top$	$\top$
$\bot$	$\bot$	$\bot$	$\top$

$\leftrightarrow_S$	$\top$	U	$\bot$
$\top$	$\top$	$\bot$	$\bot$
U	$\bot$	$\top$	$\bot$
$\bot$	$\bot$	$\bot$	$\top$

$\leftarrow_K$	$\top$	U	$\bot$
$\top$	$\top$	$\top$	$\top$
U	U	U	$\top$
$\bot$	$\bot$	U	$\top$

On the other hand, for the implication $\leftarrow_K$ we find that $I(A \vee \neg B) = I(A \leftarrow_K B) = \mathsf{U}$.

A *model* of a formula F is an interpretation I such that $I(F) = \top$. A *model* of a set of formulas is an interpretation which is a model of each formula in the set. Whether a formula is true under the given interpretation depends on the underlying three-valued logic: I is a *(three-valued) model under Łukasiewicz logic for* $\mathcal{P}$ ($I \models_Ł \mathcal{P}$) if and only if each clause occurring in $\mathcal{P}$ is mapped to $\top$ using the truth tables for $\{\neg, \wedge, \vee, \leftarrow_Ł, \leftrightarrow_Ł\}$ depicted in Table 1. As we can see from Table 1, a model of $\mathcal{P}$ under S_3 logic is a model of $\mathcal{P}$ under Łukasiewicz logic, and vice versa.

Weak Completion Semantics

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

1 Replace all clauses with the same head $A \leftarrow body_1, \ldots, A \leftarrow body_n$ by $A \leftarrow body_1 \vee \ldots \vee body_n$.

2 For all atoms A, if $A \in \mathsf{ud}\,(\mathcal{P})$ then add $A \leftarrow \bot$.

3 Replace all occurrences of $\leftarrow$ by $\leftrightarrow$.

The resulting set of equivalences is called the *completion* of $\mathcal{P}$ (Clark, 1978). If Step 2 is omitted, then the resulting set is called the *weak completion* of $\mathcal{P}$ ($\mathsf{wc}\,\mathcal{P}$) (Hölldobler & Kencana Ramli, 2009). For instance, the weak completion of $\mathcal{P} = \{p \leftarrow q\}$ is $\mathsf{wc}\,\mathcal{P} = \{p \leftrightarrow q\}$. Consequently, the three interpretations $\langle\{p, q\}, \emptyset\rangle$, $\langle\emptyset, \emptyset\rangle$ and $\langle\emptyset, \{p, q\}\rangle$ are models for $\mathsf{wc}\,\mathcal{P}$ under Łukasiewicz logic. But how to know which model is the intended one?

In Computational Logic this model is often the least model, which in many cases can be computed as least fixed points of an appropriate semantic operator (Apt & van Emden, 1982). Stenning and 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

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

As shown in Hölldobler and Kencana Ramli (2009) the least fixed point of $\Phi_{\mathcal{P}}$ is identical to the least model of the weak completion of $\mathcal{P}$ ($\mathsf{lm}_Ł\mathsf{wc}\,\mathcal{P}$). Starting with the empty interpretation $I = \langle\emptyset, \emptyset\rangle$, $\mathsf{lm}_Ł\mathsf{wc}\,\mathcal{P}$ can be computed by iterating $\Phi_{\mathcal{P}}$. Furthermore, Hölldobler and Kencana Ramli showed that the model intersection property holds for weakly completed programs. This guarantees the existence of a least model for every program.

Well-founded Semantics

Well-founded semantics is a widely accepted approach in the field of non-monotonic reasoning which has been introduced by Van Gelder et al. (1991). As shown by Przymusinski (1990), the well-founded model coincides with the least partial stable model. Partial stable model semantics (Przymusinski, 1990) is an extension of stable model semantics (Gelfond & Lifschitz, 1988) to three-valued interpretations. Stable model and partial stable semantics are only defined for normal logic programs $\mathcal{P}^+$.

Considering the least model of the weak completion of $\mathcal{P}$ ($\mathsf{lm}_Ł\mathsf{wc}\,\mathcal{P}$) and the well-founded model of $\mathcal{P}^+$ ($\mathsf{wfm}\,\mathcal{P}^+$), we observe that undefined atoms and atoms involved in positive cycles in $\mathcal{P}$ are unknown in $\mathsf{lm}_Ł\mathsf{wc}\,\mathcal{P}$, whereas in $\mathsf{wfm}\,\mathcal{P}$ they are false. However, when atoms are involved in a negative cycle in $\mathcal{P}$ they stay unknown in both $\mathsf{lm}_Ł\mathsf{wc}\,\mathcal{P}$ and $\mathsf{wfm}\,\mathcal{P}^+$.

Without loss of generality, we consider only programs where negative facts are only formulated when A is not the head of any other clause in $\mathcal{P}$. Under weak completion semantics this does not restrict the expressiveness of programs as we can only conclude that A is in $I^\bot$ if *for all* clauses where A is the head of, the body is in $I^\bot$. Thus, $A \leftarrow \bot$ would not add any more information when there is another clause with A in its head for which the body is not in $I^\bot$.

Theorem 1 (Dietz et al. (2013)) *For every tight logic program $\mathcal{P}$ and interpretation I the following two statements are equivalent:*

1 I is the least model of the weak completion of $\mathcal{P}$.

2 I is the well-founded model of $\mathcal{P}^{mod}$, where

$$\mathcal{P}^{mod} = \mathcal{P}^+ \cup \{A \leftarrow \neg n\_A,\ n\_A \leftarrow \neg A \mid A \in \mathsf{ud}\,(\mathcal{P})\}.$$

and for each $A \in \mathsf{ud}\,(\mathcal{P})$, n_A is a new atom.

The programs we discussed in Dietz et al. (2012) to model the suppression task and the programs we will discuss in the following to model the two cases of the selection task, are acyclic and thus tight. Therefore, our results hold for both of them: programs under weak completion semantics and modified programs under well-founded semantics.

The Selection Task

In the original selection task (Wason, 1968) participants were given the conditional

If there is a D on one side of the card,
then there is 3 on the other side

Table 2: The results of the abstract case of the selection task.

D	F	3	7
89%	16%	62%	25%

Table 3: The results of the social case of the selection task.

beer	coke	22 years old	16 years old
95%	0.025%	0.025%	80%

and four cards on a table showing the letters D and F as well as the numbers 3 and 7. Furthermore, they know that each card has a letter on one side and a number on the other side. Which cards must be turned over to prove that the conditional holds? Assume the conditional is represented in classical propositional logic by the implication

$$3 \leftarrow D, \tag{2}$$

where the propositional variable 3 represents the fact that the number 3 is shown and D represents the fact that the letter D is shown. Then, in order to verify the implication one must turn over the cards showing D and 7. However, as repeated experiments have shown consistently (see Table 2), participants believe differently. Whereas 89% of the participants correctly conclude that the card showing D must be turned (a number other than 3 on the other side would falsify the implication), 62% of the participants incorrectly suggest to turn over the card showing 3 (no relevant information can be found which would falsify the implication). Likewise, whereas 25% of the participants correctly believe that the card showing 7 needs to be turned over (if the other side would show a D, then the implication is falsified), 16% incorrectly believe that the card showing F needs to be turned over (no relevant information can be found which would falsify the implication). In other words, the overall correctness of the answers given for the abstract selection task if modeled by an implication in classical two-valued logic is pretty bad.

Griggs and Cox (1982) adapted Wason's selection task to a social case. Consider the conditional

If a person is drinking beer,
then the person must be over 19 years of age

and again consider four cards, where on one side there is the person's age and on the other side of the card is written what the person is drinking: *drinking beer*, *drinking coke*, *22 years old* and *16 years old*. Which drinks and persons must be checked to prove that the conditional holds? If the conditional is represented by the implication

$$o \leftarrow b, \tag{3}$$

where o represents a person being older than 19 years and b represents the person drinking beer. In order to verify the implication one must turn over the cards *drinking beer* and *16 years old*. Participants usually solve the social case of the selection task quite correctly in accordance with the laws of classical logic. Table 3 shows the results represented in Griggs and Cox (1982) for the social case. Why are the results of these two experiments so different?

Table 4: The computational logic approach for the social case of the selection task.

Card	$\mathcal{P}$	$\mathsf{lm}_{\L}\mathsf{wc}\,\mathcal{P}/\mathsf{wfm}\,\mathcal{P}^{mod}$	Griggs & Cox
beer	$\{ab_2 \leftarrow \bot, b \leftarrow \top\}$	$\langle\{b\},\{ab_2\}\rangle \not\models_{\L} (5)$	95%
coke	$\{ab_2 \leftarrow \bot, b \leftarrow \bot\}$	$\langle\emptyset,\{b, ab_2\}\rangle \models_{\L} (5)$	0.025%
16 years	$\{ab_2 \leftarrow \bot, o \leftarrow \bot\}$	$\langle\emptyset,\{o, ab_2\}\rangle \not\models_{\L} (5)$	80%
22 years	$\{ab_2 \leftarrow \bot, o \leftarrow \top\}$	$\langle\{o\},\{ab_2\}\rangle \models_{\L} (5)$	0.025%

Several attempts were made to explain these differences. Wason (1968) proposed a *defective truth table* to explain how humans reason with conditionals. When the antecedent of a conditional is false, then normally people consider the whole conditional as irrelevant and ignore it in further reasoning. Evans (1972) describes a phenomenon called the *matching bias*, where people tend to consider only the present values in the conditional. For instance, in the abstract case, card D is the easiest one to solve, because this rule is only true when both values present in the rule are on the card. On the other hand, card 7 is the most difficult one, because people have to make a double mismatch, that is, they have to consider the situation where 3 is not on the card and therefore something different than D has to be on the other side. Why do people not make these mistakes in the social case?

One explanation can be found in Kowalski (2011), namely that people view the conditional in the abstract case as a *belief*. For instance, the participants perceive the task to examine whether the rule is either true or false. On the other hand, in the social case, the participants perceive the rule as a *social constraint*, a conditional that *ought to be* true. People intuitively aim at preventing the violation of such a constraint, which is normally done by observing whether the state of the world complies with the rule. We adopt this view and model our formalism accordingly.

Modeling the Abstract and the Social Case

As already mentioned in the introduction, Stenning and van Lambalgen distinguish between two steps when modeling human reasoning. We adopt the first step, in particular, the idea to implement conditionals by licenses for implications. This can be achieved by adding an *abnormality predicate* to the antecedent of the implication. Applying this idea to the Wason selection task we obtain

$$3 \leftarrow D \wedge \neg ab_1 \tag{4}$$

instead of (2) and

$$o \leftarrow b \wedge \neg ab_2 \tag{5}$$

instead of (3), where $\neg ab_1$ and $\neg ab_2$ are used to express that the corresponding rules hold unless there are some abnormalities.

The Social Case

In this case most humans are quite familiar with the conditional as it is a standard law. They are also aware – it is common sense knowledge – that there are no exceptions or abnormalities and, hence, ab_2 is set to $\bot$.

Let us assume that conditional (5) is viewed as a social constraint which must follow logically from the given facts. Now consider the four different cases: One should observe that for the card *16 years old* the least model of the weak completion of $\mathcal{P}$, i.e. $\langle\emptyset,\{o, ab_2\}\rangle$, assigns U to b and, consequently, to both, $b \wedge \neg ab_2$ and (5), as well. Overall, for the cards *drinking beer* and *16 years old* the social constraint (5) is not entailed by the least model of the weak completion of the program. Hence, we need to turn over these cards and, hopefully, find that the beer drinker is older than 19 and that the 16 years old is not drinking beer. The results of the social case are shown in Table 4, where the last column shows the experimental results of Griggs and Cox (1982). The results of our approach correspond to the results of how the majority of the participants responded and, therefore, appears to be adequate.

The Abstract Case

This case is artificial, and consequently, there is no common sense knowledge about the conditional. Following Kowalski (2011), let us assume that conditional (4) is viewed as a belief. As there are no known abnormalities, ab_1 is set to $\bot$. Furthermore, let D, F, 3, and 7 be propositional variables denoting that the corresponding symbol or number is on one side. Altogether, we obtain the program

$$\mathcal{P} = \{3 \leftarrow D \wedge \neg ab_1,\ ab_1 \leftarrow \bot\}.$$

Its weak completion is

$$\mathsf{wc}\,\mathcal{P} = \{3 \leftrightarrow D \wedge \neg ab_1,\ ab_1 \leftrightarrow \bot\}$$

and admits the least model

$$\langle \emptyset, \{ab_1\}\rangle$$

under weak completion semantics as well as under well-founded semantics. Unfortunately, this least model does not explain any symbol on any card. We need to extend the program based on which card we observe. In order to explain an observed card, we apply abduction.

In the following we will explain abduction in the context of weak completion semantics. For tight logic programs, identical results are obtained using well-founded semantics (see Dietz et al. (2013)).

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 atom in $\mathcal{P}$ and the logical consequence relation $\models_{\text{Ł}}^{\mathsf{lm\,wc}}$, where $\mathcal{P} \models_{\text{Ł}}^{\mathsf{lm\,wc}} F$ if and only if $\mathsf{lm}_{\text{Ł}}\mathsf{wc}\,\mathcal{P}(F) = \top$ for the formula F. As *observations* we consider literals.

Let $\langle \mathcal{P}, \mathcal{A}, \models_{\text{Ł}}^{\mathsf{lm\,wc}}\rangle$ be an abductive framework and $\mathcal{O}$ an observation. $\mathcal{O}$ is *explained* by $\mathcal{E}$ if and only if $\mathcal{E} \subseteq \mathcal{A}$, $\mathcal{P} \cup \mathcal{E}$ is satisfiable, and $\mathcal{P} \cup \mathcal{E} \models_{\text{Ł}}^{\mathsf{lm\,wc}} \mathcal{O}$. Usually, minimal explanations are preferred. In case there exist several minimal explanations, then two forms of reasoning can be distinguished. F follows *skeptically* from program $\mathcal{P}$ and observation $\mathcal{O}$ if and only if $\mathcal{O}$ can be explained and for all minimal explanations $\mathcal{E}$ we find $\mathcal{P} \cup \mathcal{E} \models_{\text{Ł}}^{\mathsf{lm\,wc}} \mathcal{O}$, whereas F follows *credulously* from $\mathcal{P}$ and $\mathcal{O}$ if and only if there exists a minimal explanation $\mathcal{E}$ such that $\mathcal{P} \cup \mathcal{E} \models_{\text{Ł}}^{\mathsf{lm\,wc}} \mathcal{O}$.

In the case of the abstract case of the Wason selection task, the set of abducibles is

$$\{D \leftarrow \top,\ D \leftarrow \bot,\ F \leftarrow \top,\ F \leftarrow \bot,\ 7 \leftarrow \top,\ 7 \leftarrow \bot\}.$$

Now consider the four different cases, where the explanations $\mathcal{E}$ are minimal. In the cases where F or 7 were observed, the least model of the weak completion of $\mathcal{P} \cup \mathcal{E}$ does not contain any information that needs to be verified and simply confirms the observation; no further action is needed. In some sense, the belief about the premises and conclusions of the conditional is irrelevant. The truth values of them are unknown and under Łukasiewicz logic this makes the conditional true.

Table 5: The computational logic approach for the abstract case of the selection task.

$\mathcal{O}$	$\mathcal{E}$	$\mathsf{lm}_{\text{Ł}}\mathsf{wc}\,(\mathcal{P} \cup \mathcal{E})/\mathsf{wfm}\,(\mathcal{P} \cup \mathcal{E})^{mod}$		Wason
D	$\{D \leftarrow \top\}$	$\langle\{D,\ 3\}, \{ab_1\}\rangle$	$\rightsquigarrow$ turn over	89%
F	$\{F \leftarrow \top\}$	$\langle\{F\}, \{ab_1\}\rangle$	$\rightsquigarrow$ no turn over	16%
3	$\{D \leftarrow \top\}$	$\langle\{D,\ 3\}, \{ab_1\}\rangle$	$\rightsquigarrow$ turn over	62%
7	$\{\ 7 \leftarrow \top\}$	$\langle\{\ 7\}, \{ab_1\}\rangle$	$\rightsquigarrow$ no turn over	25%

In the case where D was observed, the least model maps also 3 to $\top$. That means, in order to be sure that this corresponds to the real situation, we need to check if 3 is true. Therefore, the card showing D is turned over. Likewise, in the case where 3 is observed, D is also mapped to $\top$ in the least model, which can only be confirmed if the card is turned over. As in each case there is a single minimal explanation, there is no need to distinguish between sceptical and credulous reasoning. The results of the abstract case are shown in Table 5, where the last column shows the experimental results of Wason (1968). The results of our approach correspond to the results of how the majority of the participants responded and, therefore, appears to be adequate.

A Psychological Study

One of the main differences between weak completion and well-founded semantics is how they deal with positive cycles in logic programs. While in a well-founded model atoms involved in positive cycles are false, they are mapped to unknown under the weak completion semantics. In order to determine which semantics is more adequate for human reasoning, we need to investigate which conclusions are typically drawn by human reasoners with respect to cyclic conditionals. For this purpose we carried out a psychological study.

Participants We tested 35 participants on an online website (Amazon Mechanical Turk). They were paid for their participation.

Material, Procedure and Design Participants were presented with 17 problems consisting of cyclic conditionals of length 1, 2 and 3. Consider the following cyclic conditional of length 1:

If they open the window, then they open the window.

Participants were asked about the consequences of this conditional and could choose between one of the following three offered conclusions: *They open the window*, *they do no open the window*, and *it is unknown whether they open the window.* Another example is the following cyclic conditional of length 3:

If they open the window, then it is cold.
If it is cold, then they wear their jackets.
If they wear their jackets, then they open the window.

We investigated three kinds of atoms, viz. whether they open the window, whether it is cold, and whether they wear their

Table 6: The length of the cycles, the given answers, and their mean response times.

Length	Chosen answer in percentage			Mean response
of cycle	Positive	Negative	Unknown	times in msec
1	75	0	25	5267
2	60	3	37	11516
3	55	4	41	11680

jacket; each of them under the three conditions positive, negative, and unknown.

Results and Discussion The results (cf. Table 6) indicate two kinds of groups each taking a different interpretation of the statements: One group consists of participants understanding the programs as a conditional, which in our approach is modeled by $p \leftarrow p \wedge \neg ab$ for cycles of length one ($p \leftarrow q \wedge \neg ab_1, q \leftarrow p \wedge \neg ab_2$ for cycles of length 2, and accordingly for cycles of length 3). If we assume that nothing abnormal is known, (i.e., $ab \leftarrow \bot$), then the least model of the weak completion is $\langle \emptyset, \{ab\} \rangle$. In contrast, the well-founded semantics always and independently of the truth value of ab concludes $\neg p$, a conclusion almost no participant has drawn. The other interpretation, where participants' chose to give a positive answer, apparently treats the statement as a fact $p \leftarrow \top$. If we consider this as the result of the first step of the Stenning and van Lambalgen procedure (reasoning towards an adequate representation) then both, weak completion and well-founded semantics seem to be adequate. The findings show that the chosen answers associated with positive atoms decrease from cycles of length 1 (75% positive answers) to cycles of length 3 (55% positive answers) with an increase of choosing the truth-value unknown. The response times indicate a higher degree of uncertainty in case of problems involving cycles of length 2 and 3 in contrast to the simpler problems involving a cycle of length 1. Taken together, the increase in choosing the truth value unknown and the increase in response time shows an increasing likelihood of the participants to adopt a weak completion semantics.

When participants were given conditionals with negative cycles of the form $p \leftarrow \neg q \wedge \neg ab_1, q \leftarrow \neg p \wedge \neg ab_2$, then the majority concluded that the given facts were unknown. This result corresponds to both, weak completion and well-founded semantics.

Summing up, it seems that, when we consider the two representational forms for the conditionals, then weak completion semantics can better explain and predict participants' responses than well-founded semantics. As discussed in (Wernhard, 2012), it would be interesting to further examine whether there are real world situations in which humans actually reason with cycles and how they extract knowledge based on these seemingly meaningless data.

Conclusion

We have presented a computational logic approach for modeling human reasoning in the Wason selection task. It is based on a previously proposed approach that adequately models another psychological study, the suppression task. We extended our approach following an idea from Kowalski's task representation: in order to solve the social case correctly, the conditional must be seen as a social constraint, whereas the abstract case is correctly represented when the conditional is seen as a belief. show The second case can be modeled by extending the formalization to reasoning (either credulously or sceptically) within an abductive framework. Hölldobler, Philipp, and Wernhard (2011) have shown that sceptical reasoning has to be applied to solve the suppression task.

Stenning and van Lambalgen analyzed the Wason selection task but did not attempt to formalize this task based on their previous approach for the suppression task. On the other hand, Kowalski showed how to formalize the abstract and the social case of the selection task, but did not propose a solution to the suppression task. In our paper, we present one approach which seems to adequately model both tasks.

However, there are still aspects we did not consider yet and which need to be further examined. Our approach does not deal with the so-called first step of modeling human reasoning: reasoning with respect to an adequate representation. We just assume that in the social case people take the conditional as a social constraint whereas they take it as a belief in the abstract case. These differences are modeled outside of the formal framework.

Dawson and Regan (2002) show by some psychological experiments that the so-called *confirmation bias* plays an important role in the Wason selection task: if people disagree with the statement of the conditional, they are more likely to find the solution because they are motivated to search for a counterexample which refutes the conditional. On the other hand, people who agree with the statement of the conditional take it as a confirmation of their believes and therefore will not extensively search to falsify the conditional.

An interesting observation discussed in Stenning and Lambalgen (2008) is that similar to the *verification bias*, people might transfer the *truth of the card* to the *truth of the rule*. In the social case, this confusion cannot occur, because it is commonsense that the rule is true, independent from whether people behave accordingly. This leads to another phenomenon, namely that participants see a dependency between the card choices and might prefer to solve the problem by *reactive planning*. They would only like to decide what to do after they saw the outcome of the first card. For instance, if one turns over card D first and there is no 3 on the other side, no further cards needs to be examined, because the rule has been falsified. However, if there is a 3 on the other side, the other options need to be considered again. This kind of behavior could be described in a framework with belief change: Each card which is turned over is a piece of new information which needs to be integrated into the current knowledge base and updates new inferences accordingly.

Acknowledgments

The authors are grateful to Bertram Fronhöfer, Luís Moniz Pereira and Christoph Wernhard for many helpful discussions.

References

Apt, K. R., Blair, H. A., & Walker, A. (1988). Foundations of deductive databases and logic programming. In J. Minker (Ed.), *Towards a theory of declarative knowledge* (pp. 89–148). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Apt, K. R., & van Emden, M. H. (1982). Contributions to the theory of logic programming. *Journal of the ACM*, *29*, 841–862.

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

Clark, K. L. (1978). Negation as failure. In H. Gallaire & J. Minker (Eds.), *Logic and data bases* (Vol. 1, pp. 293–322). New York, London: Plenum Press.

Dawson, T., E. Gilovich, & Regan, D. T. (2002). Motivated reasoning and performance on the wason selection task. *Personality and Social Psychology Bulletin*, *28*, 1379–1387.

Dietz, E.-A., Hölldobler, S., & Ragni, M. (2012). A computational logic approach to the suppression task. In D. P. N. Miyake & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1500–1505). Cognitive Science Society.

Dietz, E.-A., Hölldobler, S., & Wernhard, C. (2013). *Modeling the Suppression Task under Weak Completion and Well-Founded Semantics* (Tech. Rep. No. KRR-13-02). Technische Universität Dresden. (submitted)

Erdem, E., & Lifschitz, V. (2003). Tight logic programs. *Theory and Practice of Logic Programming*, *3*(4), 499–518.

Evans, J. (1972). Interpretation and matching bias in a reasoning task. *British Journal of Psychology*, *24*, 193–199.

Fitting, M. (1985). A Kripke-Kleene semantics for logic programs. *The Journal of Logic Programming*, *2*(4).

Gelfond, M., & Lifschitz, V. (1988). The stable model semantics for logic programming. In R. A. Kowalski & K. Bowen (Eds.), *Proceedings of international logic programming conference and symposium* (pp. 1070–1080). MIT Press.

Gottwald, S. (2001). *A treatise on many-valued logics* (Vol. 9). Baldock, UK: Research Studies Press.

Griggs, R., & Cox, J. (1982). The elusive thematic materials effect in the Wason selection task. *British Journal of Psychology*, *73*, 407–420.

Hölldobler, S., & Kencana Ramli, C. D. (2009). Logic Programs under Three-Valued Łukasiewicz Semantics. In P. M. Hill & D. S. Warren (Eds.), *International conference on logic programming, LNCS* (Vol. 5649, pp. 464–478). Berlin, Heidelberg: Springer-Verlag.

Hölldobler, S., & Kencana Ramli, C. D. (2009). Logics and networks for human reasoning. In C. Alippi, M. M. Polycarpou, C. G. Panayiotou, & G. Ellinas (Eds.), *International Conference on Artificial Neural Networks 2009, Part II, LNCS* (Vol. 5769, pp. 85–94). Berlin, Heidelberg: Springer-Verlag.

Hölldobler, S., Philipp, T., & Wernhard, C. (2011). An abductive model for human reasoning. In *Logical formalizations of commonsense reasoning, papers from the AAAI 2011 spring symposium* (pp. 135–138). AAAI Press.

Kakas, A. C., Kowalski, R. A., & Toni, F. (1993). Abductive logic programming. *Journal of Logic and Computation*, *2*(6), 719–770.

Kleene, S. C. (1952). *Introduction to metamathematics*. Amsterdam: North-Holland.

Kowalski, R. (2011). *Computational logic and human thinking: How to be artificially intelligent*. Cambridge: Cambridge University Press.

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, Amsterdam, 87-88, 1990.)

Przymusinski, T. (1989). Every logic program has a natural stratification and an iterated least fixed point model. In *Proceedings of the eighth ACM SIGACT-SIGMOD-SIGART symposium on principles of database systems* (pp. 11–21). New York, NY, USA: ACM.

Przymusinski, T. (1990). Well-founded semantics coincides with three-valued stable semantics. *Fundamenta Informaticae*, *13*(4), 445–463.

Rescher, N. (1969). *Many-valued logic*. New York: McGraw-Hill.

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

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

Van Gelder, A., Ross, K. A., & Schlipf, J. S. (1991). The well-founded semantics for general logic programs. *Journal of the ACM*, *38*(3), 619–649.

Wason, P. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*(3), 273–281.

Wernhard, C. (2012). Towards a declarative approach to model human reasoning with nonmonotonic logics. In T. Barkowsky, M. Ragni, & F. Stolzenburg (Eds.), *Human reasoning and automated deduction: KI 2012 workshop proceedings* (Vol. SFB/TR 8 Report 032-09/2012, pp. 41–48). Universität Bremen / Universität Freiburg, Germany.