A Computational Logic Approach to Syllogisms in Human Reasoning

Emmanuelle-Anna Dietz
dietz@iccl.tu-dresden.de

International Center for Computational Logic, TU Dresden, 01062 Dresden, Germany

Abstract. Psychological experiments on syllogistic reasoning have shown that participants did not always deduce the classical logically valid conclusions. In particular, the results show that they had difficulties to reason with syllogistic statements that contradicted their own beliefs. This paper discusses syllogisms in human reasoning and proposes a formalization under the weak completion semantics.

1 Introduction

Evans, Barston and Pollard [10] made a psychological study about deductive reasoning, which demonstrated possibly conflicting processes in human reasoning. Participants were presented different syllogisms, for which they had to decide whether these were (classical) logically valid. Consider S_{vit}:

Premise 1	*No nutritional things are inexpensive.*
Premise 2	*Some vitamin tablets are inexpensive.*
Conclusion	*Therefore, some vitamin tablets are not nutritional.*

The conclusion necessarily follows from the premises. However, approximately half of the participants said that this syllogism was not logically valid. They were explicitly asked to logically validate or invalidate various syllogisms. Table 1 gives four examples of syllogisms, which have been tested in [10]. If participants judged that "the conclusion necessarily follows from the statements in the passage, [you]" they "should answer 'yes,' otherwise 'no'." The last column shows the percentage of the participants that believed the syllogism to be valid. Evans, Barston and Pollard asserted that the participants were influenced by their own beliefs, their so-called belief bias, where we distinguish between the negative and the positive belief bias [11]. The negative belief bias, i.e., when a support for the unbelievable conclusion is suppressed, happens for 56% of the participants in S_{vit}. A positive belief bias, i.e., when the acceptance for the believable conclusion is raised, happens for 71% of the participants in S_{cig}. As pointed out in [14], Wilkins [32] already observed that syllogisms, which conflict with our beliefs are more difficult to solve. People reflectively read the instructions and understand well that they are required to reason logically from the premises to the conclusion. However, the results show that their intuitions are stronger and deliver a tendency to say 'yes' or 'no' depending on whether it

	Type	Case	%
S_{dog}	valid and believable	*No police dogs are vicious.* *Some highly trained dogs are vicious.* *Therefore, some highly trained dogs are not police dogs.*	89
S_{vit}	valid and unbelievable	*No nutritional things are inexpensive.* *Some vitamin tablets are inexpensive.* *Therefore, some vitamin tablets are not nutritional.*	56
S_{rich}	invalid and unbelievable	*No millionaires are hard workers.* *Some rich people are hard workers.* *Therefore, some millionaires are not rich people.*	10
S_{cig}	invalid and believable	*No addictive things are inexpensive.* *Some cigarettes are inexpensive.* *Therefore, some addictive things are not cigarettes.*	71

Table 1. Examples of four kinds of syllogisms. The percentages are summarized results over three experiments and show the rate that the conclusion is accepted to be valid [10].

is believable [9]. Various theories have tried to explain this phenomenon. Some conclusions can be explained by converting the premises [2] or by assuming that the atmosphere of the premises influences the acceptance for the conclusion [33]. Johnson-Laird and Byrne [20] proposed the mental model theory [19], which additionally supposes the search for counterexamples when validating the conclusion. These theories have been partly rejected or claimed to be incomplete. Evans et al. [10, 12] proposed a theory, which is sometimes referred to as the selective scrutiny model [1, 14]. First, humans heuristically accept any syllogism having a believable conclusion, and only check on the logic if the conclusion contradicts their belief. Adler and Rips [1] claim that this behavior is rational because it efficiently maintains our beliefs, except in case if there is any evidence to change them. It results in an adaptive process, for which we only make an effort towards a logical evaluation when the conclusion is unbelievable. It would take a lot of effort if we would constantly verify them even though there is no reason to question them. As people intend to keep their beliefs as consistent as possible, they invest more effort in examining statements that contradict them, than the ones that comply with them. However, this theory cannot fully explain all classical logical errors in the reasoning process. Yet another approach, the selective processing model [8], accounts only for a single preferred model. If the conclusion is neutral or believable, humans attempt to construct a model that supports it. Otherwise, they attempt to construct a model, which rejects it.

As summarized in [14], there are several stages in which a belief bias can take place. First, beliefs can influence our interpretation of the premises. Second, in case a statement contradicts our belief, we might search for alternative models and check whether the conclusion is plausible.

Stenning and van Lambalgen [30] explain why certain aspects influence the interpretations made by humans when evaluating syllogisms and discuss this in the context of mental models. They propose to model human reasoning in a

two step procedure. First, human reasoning should be modeled towards an adequate representation. Second, human reasoning should be adequately modeled with respect to this representation. In our context, the first step is about the representational part, that is, which our beliefs influence the interpretation of the premises. The second step is about the procedural part, that is, whether we search for alternative models and whether the conclusion is plausible.
After we have specified some preliminaries, we explain in Section 3 how the just discussed four cases of the syllogistic reasoning task can be represented in logic programs. Based on this representation, Section 4 discusses how beliefs and background knowledge influences the reasoning process and shows that the results can be modeled by computing the least models of the weak completion.

2 Preliminaries

The general notation, which we will use in the paper, is based on [15, 22].

2.1 Logic Programs

We restrict ourselves to datalog programs, i.e., the set of terms consists only of constants and variables. A *logic program* $\mathcal{P}$ is a finite set of clauses of the form

$$A \leftarrow L_1 \wedge \ldots \wedge L_n, \tag{1}$$

where $n \geq 0$ with finite n. A is an atom and L_i, $1 \leq i \leq n$, are literals. A is called *head* of the clause and the subformula to the right of the implication sign is called *body* of the clause. If the clause contains variables, then they are implicitly universally quantified within the scope of the entire clause. A clause that does not contain variables, is called a *ground* clause. In case $n = 0$, the clause is a *positive fact* and denoted as

$$A \leftarrow \top.$$

A *negative fact* is denoted as

$$A \leftarrow \bot,$$

where *true* ,$\top$, and *false*, $\bot$, are *truth-value constants*. The notion of falsehood appears counterintuitive at first sight, but programs will be interpreted under their (weak) completion where we replace the implication by the equivalence sign. We assume a fixed set of constants, denoted by $\mathsf{CONSTANTS}$, which is nonempty and finite. $\mathsf{constants}(\mathcal{P})$ denotes the set of all constants occurring in $\mathcal{P}$. If not stated otherwise, we assume that $\mathsf{CONSTANTS} = \mathsf{constants}(\mathcal{P})$.
$\mathsf{g}\,\mathcal{P}$ denotes ground $\mathcal{P}$, which means that $\mathcal{P}$ contains exactly all the ground clauses with respect to the alphabet. $\mathsf{atoms}(\mathcal{P})$ denotes the set of all atoms occurring in $\mathcal{P}$. If atom A is not the head of any clause in $\mathcal{P}$, then A is *undefined* in $\mathcal{P}$. The set of all atoms that are undefined in $\mathcal{P}$, is denoted by $\mathsf{undef}(\mathcal{P})$.

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	$\bot$

$\leftarrow_{\text{Ł}}$	$\top$	U	$\bot$
$\top$	$\top$	$\top$	$\top$
U	U	$\top$	$\top$
$\bot$	$\bot$	U	$\top$

$\leftrightarrow_{\text{Ł}}$	$\top$	U	$\bot$
$\top$	$\top$	U	$\bot$
U	U	$\top$	U
$\bot$	$\bot$	U	$\top$

Table 2. $\top$, $\bot$, and U denote *true*, *false*, and *unknown*, respectively.

2.2 Three-Valued Łukasiewicz Semantics

We consider the three-valued Łukasiewicz Semantics [23], for which the corresponding truth values are $\top$, $\bot$ and U, which mean *true*, *false* and *unknown*, respectively. A *three-valued interpretation* I is a mapping from formulas to a set of truth values $\{\top, \bot, \mathsf{U}\}$. The truth value of a given formula under I is determined according to the truth tables in Table 2. 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. Atoms, which do not occur in $I^\top \cup I^\bot$, are mapped to U. Let $I = \langle I^\top, I^\bot \rangle$ and $J = \langle J^\top, J^\bot \rangle$ be two interpretations: $I \subseteq J$ iff $I^\top \subseteq J^\top$ and $I^\bot \subseteq J^\bot$. $I(F) = \top$ means that a formula F is mapped to true under I. $\mathcal{M}$ is a *model* of $\mathsf{g}\,\mathcal{P}$ if it is an interpretation, which maps each clause occurring in $\mathsf{g}\,\mathcal{P}$ to $\top$. I is the *least model* of $\mathsf{g}\,\mathcal{P}$ iff for any other model J of $\mathsf{g}\,\mathcal{P}$ it holds that $I \subseteq J$.

2.3 Reasoning with Respect to Least Models

Consider following transformation for $\mathsf{g}\,\mathcal{P}$:

1. Replace all clauses in $\mathsf{g}\,\mathcal{P}$ with the same head $A \leftarrow Body_1, A \leftarrow Body_2, \ldots$ by the single expression $A \leftarrow Body_1 \vee Body_2, \vee \ldots$.
2. If $A \in \mathsf{undef}(\mathsf{g}\,\mathcal{P})$, then add $A \leftarrow \bot$.
3. Replace all occurrences of $\leftarrow$ by $\leftrightarrow$.

The resulting set of equivalences is called the *completion* of $\mathsf{g}\,\mathcal{P}$ [3]. If Step 2 is omitted, then the resulting set is called the *weak completion* of $\mathsf{g}\,\mathcal{P}$ ($\mathsf{wc}\,\mathsf{g}\,\mathcal{P}$). In contrast to completed programs, the model intersection property holds for weakly completed programs [17]. This guarantees the existence of a least model for every program. Stenning and van Lambalgen [30] devised such an operator, which has been generalized for first-order programs by [16]: Let I be an interpretation in $\Phi_{\mathsf{SvL},\mathcal{P}}(I) = \langle J^\top, J^\bot \rangle$, where

$$J^\top = \{A \mid \text{there exists a clause } A \leftarrow Body \in \mathsf{g}\,\mathcal{P} \text{ with } I(Body) = \top\},$$
$$J^\bot = \{A \mid \text{there exists a clause } A \leftarrow Body \in \mathsf{g}\,\mathcal{P} \text{ and}$$
$$\text{for all clauses } A \leftarrow Body \in \mathsf{g}\,\mathcal{P} \text{ we find } I(Body) = \bot\}.$$

As shown in [16] the least fixed point of $\Phi_{SvL,\mathcal{P}}$ is identical to the least model of the weak completion of $\mathsf{g}\,\mathcal{P}$ under three-valued Łukasiewicz semantics. In the

following, we will denote the least model of the weak completion of a given program $\mathcal{P}$ by $\mathsf{lm}_{\text{L}}\mathsf{wc}\,\mathsf{g}\,\mathcal{P}$. From $I = \langle \emptyset, \emptyset \rangle$, $\mathsf{lm}_{\text{L}}\mathsf{wc}\,\mathsf{g}\,\mathcal{P}$ is computed by iterating $\Phi_{SvL,\mathcal{P}}$. Given a program $\mathcal{P}$ and a formula F, $\mathcal{P} \models_{\text{L}}^{\mathsf{lmwc}} F$ iff $\mathsf{lm}_{\text{L}}\mathsf{wc}\,\mathsf{g}\,\mathcal{P}(F) = \top$ for formula F. Notice that Φ_{SvL} differs in a subtle way from the well-known Fitting operator Φ_F, introduced in [13]: The definition of Φ_F is like that of Φ_{SvL}, except that in the specification of $J^\perp$ the first line "there exists a clause $A \leftarrow \mathit{Body} \in \mathsf{g}\,\mathcal{P}$ and" is dropped. The least fixed point of $\Phi_{F,\mathcal{P}}$ corresponds to the least model of the completion of $\mathsf{g}\,\mathcal{P}$. If an atom A is undefined in $\mathsf{g}\,\mathcal{P}$, then, for arbitrary interpretations I it holds that $A \in J^\perp$ in $\Phi_{F,\mathcal{P}}(I) = \langle J^\top, J^\perp \rangle$, whereas if Φ_{SvL} is applied instead of Φ_F, this does not hold for any interpretation I.
The correspondence between weak completion semantics and well-founded semantics [31] for tight programs, i.e. those without positive cycles, is shown in [6].

2.4 Integrity Constraints

A set of *integrity constraints* $\mathcal{IC}$ comprises clauses of the form $\perp \leftarrow \mathit{Body}$, where *Body* is a conjunction of literals. Under three-valued semantics, there are several ways on how to understand integrity constraints [21], two of them being the *theoremhood view* and the *consistency view.* Consider $\mathcal{IC}$:

$$\perp \leftarrow \neg p \wedge q.$$

The theoremhood view requires that a model only satisfies the set of integrity constraints if for all its clauses, *Body* is false under this model. In the example, this is only the case if p is true or if q is false in the model. In the consistency view, the set of integrity constraints is satisfied by the model if *Body* is unknown or false in it. Here, a model satisfies $\mathcal{IC}$ already if either p or q is unknown.
Given $\mathcal{P}$ and set $\mathcal{IC}$, $\mathcal{P}$ *satisfies* $\mathcal{IC}$ iff there exists I, which is a model for $\mathsf{g}\,\mathcal{P}$, and for each $\perp \leftarrow \mathit{Body} \in \mathcal{IC}$, we find that $I(\mathit{Body}) \in \{\perp, \mathsf{U}\}$.

2.5 Abduction

We extend two-valued abduction [21] for three-valued semantics. The set of abducibles $\mathcal{A}_\mathcal{P}$ may not only contain positive but can also contain negative facts:

$$\{A \leftarrow \top \mid A \in \mathsf{undef}(\mathcal{P})\} \cup \{A \leftarrow \perp \mid A \in \mathsf{undef}(\mathcal{P})\}.$$

Let $\langle \mathcal{P}, \mathcal{A}_\mathcal{P}, \mathcal{IC}, \models_{\text{L}}^{\mathsf{lmwc}} \rangle$ be an abductive framework, $\mathcal{E} \subset \mathcal{A}_\mathcal{P}$ and observation $\mathcal{O}$ a non-empty set of literals.

$\mathcal{O}$ is *explained by* $\mathcal{E}$ *given* $\mathcal{P}$ *and* $\mathcal{IC}$ iff
  $\mathcal{P} \not\models_{\text{L}}^{\mathsf{lmwc}} \mathcal{O}$, $\mathcal{P} \cup \mathcal{E} \models_{\text{L}}^{\mathsf{lmwc}} \mathcal{O}$ and $\mathsf{lm}_{\text{L}}\mathsf{wc}\,\mathsf{g}\,(\mathcal{P} \cup \mathcal{E})$ *satisfies* $\mathcal{IC}$.
$\mathcal{O}$ is *explained given* $\mathcal{P}$ *and* $\mathcal{IC}$ iff
  there exists an $\mathcal{E}$ such that $\mathcal{O}$ is explained by $\mathcal{E}$ given $\mathcal{P}$ and $\mathcal{IC}$.

We assume that explanations are minimal, that means, there is no other explanation $\mathcal{E}' \subset \mathcal{E}$ for $\mathcal{O}$. In case abducibles are not abduced as positive or negative facts, they stay unknown in the least model of the weak completion. We distinguish between skeptical and credulous abduction as follows:

F *follows skeptically from* $\mathcal{P}$, $\mathcal{IC}$ *and* $\mathcal{O}$ iff $\mathcal{O}$ can be explained given $\mathcal{P}$ and $\mathcal{IC}$, and for all minimal $\mathcal{E}$ for $\mathcal{O}$, given $\mathcal{P}$ and $\mathcal{IC}$, it holds that $\mathcal{P} \cup \mathcal{E} \models^{\mathsf{lmwc}}_{\mathrm{Ł}} F$.

F *follows credulously from* $\mathcal{P}$, $\mathcal{IC}$ *and* $\mathcal{O}$ iff there exists a minimal $\mathcal{E}$ for $\mathcal{O}$, given $\mathcal{P}$ and $\mathcal{IC}$, and it holds that $\mathcal{P} \cup \mathcal{E} \models^{\mathsf{lmwc}}_{\mathrm{Ł}} F$.

3 Reasoning Towards an Appropriate Logical Form

Let us specify the syllogisms from the introduction in logic programs. We first discuss a technical aspect that allows us to encode the negative consequences of the premises. Section 3.2 covers the representational part and show how the beliefs, which might influence the interpretation of the premises, are encoded.

3.1 Positive Encoding of Negative Consequences

The first premise of S_{dog} is

No police dogs are vicious.

and is equivalent to

If something is vicious, then it is not a police dog.

and *If something is a police dog, then it is not vicious.*

The consequences in both inferences are the negation of *it is a police dog* and the negation of *it is vicious*, respectively. As the weak completion semantics does not allow negative heads in clauses, we cannot represent these inferences in a logic program straightaway. For every negative conclusion $\neg p(X)$ we introduce an auxiliary formula $p'(X)$ together with the clause $p(X) \leftarrow \neg p'(X)$. We obtain the following preliminary representation of the first premise of S_{dog} wrt *vicious*:[1]

$$police\_dog'(X) \leftarrow vicious(X), \qquad police\_dog(X) \leftarrow \neg police\_dog'(X),$$

where $police\_dog(X)$, $police\_dog'(X)$, and $vicious(X)$ denote that X is a police dog, X is not a police dog, and X is vicious, respectively. A model $I = \langle I^\top, I^\bot \rangle$ that contains both $police\_dog(X)$ and $police\_dog'(X)$ in $I^\top$ should be invalidated. This condition can be represented by the integrity constraint

$$\mathcal{IC}_{police\_dog} = \{\bot \leftarrow police\_dog(X) \wedge police\_dog'(X)\},$$

and is to be understood as discussed in Section 2.4. For the following examples, whenever there exists a $p(X)$ and its $p'(X)$ counterpart in $\mathcal{P}$, we implicitly assume $\mathcal{IC}_p = \{\bot \leftarrow p(X) \wedge p'(X)\}$.

[1] In the following we will only encode one of the inferences.

3.2 Abnormality Predicates and Background Knowledge

Newstead and Griggs [25] have shown, that the universal quantifiers in natural language are often understood as fuzzy quantifiers, which allow exceptions. In some circumstances *for all* is understood as *for almost all*. They argue that the statement *all Germans are hardworking* seems to permit exceptions and is understood as a generalization about all Germans and not a statement, which is true for each one.
This fuzzy interpretation of quantifiers seems to be in line with Stenning and van Lambalgen's suggestion to implement conditionals by default licenses for implications [29, 30]. They propose to introduce abnormality predicates, which should be added to the antecedent of the implication, where the abnormality predicate is initially assumed to be false. Consider again PREMISE 1 in S_{dog}, which can be understood as

If something is vicious and not abnormal (in that respect),
then it is not a police dog.
Nothing (by default) is abnormal (regarding the previous sentence).

This information together with the previously introduced clauses for PREMISE 1 in S_{dog} can now be encoded as:

$$\begin{aligned} police\_dog'(X) &\leftarrow vicious(X) \wedge \neg ab_{dog'}(X), \\ police\_dog(X) &\leftarrow \neg police\_dog'(X), \\ ab_{dog'}(X) &\leftarrow \bot. \end{aligned}$$

$\mathbf{S}_{dog}$ PREMISE 2 states that there are some highly trained dogs that are vicious. This statement presupposes that there actually exists something, let us say a new reserved (Skolem) constant a, for which the following is true:

$$highly\_trained(a) \leftarrow \top \qquad \text{and} \qquad vicious(a) \leftarrow \top.$$

$\mathcal{P}_{dog}$ represents the first two premises of S_{dog}:

$$\begin{aligned} police\_dog'(X) &\leftarrow vicious(X) \wedge \neg ab_{dog'}(X), \\ police\_dog(X) &\leftarrow \neg police\_dog'(X), \\ ab_{dog'}(X) &\leftarrow \bot, \\ highly\_trained(a) &\leftarrow \top, \\ vicious(a) &\leftarrow \top. \end{aligned}$$

We encode the first two premises of the other syllogisms similarly.

$\mathbf{S}_{vit}$ PREMISE 2 states that there are some vitamin tablets, which are inexpensive. We presuppose that there exists something, a, for which these facts are true:

$$vitamin(a) \leftarrow \top \qquad \text{and} \qquad inex(a) \leftarrow \top.$$

Additionally, it is commonly known that

The purpose of vitamin tablets is to aid nutrition.

This belief and the clause representing PREMISE 1 leads to

If something is a vitamin tablet, then it is abnormal (regarding PREMISE 1 *of* S_{vit}*).*

The program $\mathcal{P}_{vit}$ represents PREMISE 1 and PREMISE 2 together with the background knowledge:

$$
\begin{aligned}
nutritional'(X) &\leftarrow inex(X) \wedge \neg ab(X),\\
nutritional(X) &\leftarrow \neg nutritional'(X),\\
ab(X) &\leftarrow \bot,\\
ab(X) &\leftarrow vitamin(X),\\
vitamin(a) &\leftarrow \top,\\
inex(a) &\leftarrow \top.
\end{aligned}
$$

$nutritional(X)$, $nutritional'(X)$ denote X is nutritional, not nutritional, resp.

$\mathbf{S}_{\boldsymbol{rich}}$ PREMISE 2 states that there are some hard workers who are rich. We presuppose that there is someone, let us say, a, for which these facts are true:

$$hard\_worker(a) \leftarrow \top \qquad \text{and} \qquad rich(a) \leftarrow \top.$$

$\mathcal{P}_{rich}$ represents PREMISE 1 and PREMISE 2 of S_{rich}:

$$
\begin{aligned}
mil'(X) &\leftarrow hard\_worker(X) \wedge \neg ab(X),\\
mil(X) &\leftarrow \neg mil'(X),\\
ab(X) &\leftarrow \bot,\\
rich(a) &\leftarrow \top,\\
hard\_worker(a) &\leftarrow \top.
\end{aligned}
$$

$mil(X)$ and $mil'(X)$ denote X is a millionaire and not a millionaire, resp.

$\mathbf{S}_{\boldsymbol{cig}}$ PREMISE 2 states that there are some cigarettes, which are inexpensive. Again, we presuppose that there is something, a, for which these facts are true:

$$cig(a) \leftarrow \top \qquad \text{and} \qquad inex(a) \leftarrow \top.$$

Additionally, it is commonly known that

Cigarettes are addictive.

This belief and the clause representing PREMISE 1 leads to

If something is a cigarette, then it is abnormal
(regarding PREMISE 1 *of S_{cig}).*

As discussed by Evans et al. [10], humans seem to have a background knowledge or belief, which might provide the motivation on whether to validate a syllogism. A direct representation of PREMISE 2 is

There exists a cigarette, which is inexpensive. (1)

Additionally, in the context of PREMISE 1, we assume that

Compared to other addictive things, cigarettes are inexpensive. (2)

which implies (1) and biases the reasoning towards a representation. Note that (2) only implies (1) because we understand quantifiers with existential import, i.e., *for all* implies *there exists*. This is a reasonable assumption when modeling human reasoning, as in natural language we normally do not quantify over things that don't exist. Furthermore, Stenning and an Lambalgen [30] have shown that humans require existential import for the conditional to be true.
The belief bias represented by (2), together with the idea to represent conditionals by a normal default permission for implication, leads to the conditional

If something is a cigarette and not abnormal, then it is inexpensive. (3)
Nothing (as a rule) is abnormal (regarding (3)*).*

$\mathcal{P}_{cig}$ represents the first two premises and the background knowledge in S_{cig} as follows:

$$\begin{aligned} addictive'(X) &\leftarrow inex(X) \wedge \neg ab_{add'}(X), \\ addictive(X) &\leftarrow \neg addictive'(X), \\ ab_{add'}(X) &\leftarrow \bot, \\ ab_{add'}(X) &\leftarrow cig(X), \\ inex(X) &\leftarrow cig(X) \wedge \neg ab_{inex}(X), \\ ab_{inex}(X) &\leftarrow \bot, \\ cig(a) &\leftarrow \top, \\ inex(a) &\leftarrow \top, \end{aligned}$$

$addictive(X)$ and $addictive'(X)$ denote X is addictive and not addictive, resp.

4 Reasoning with Respect to Least Models

This section deals with Stenning and van Lambalgen's second step, and discusses where a possible belief bias during the reasoning procedure can influence the result. We show how to compute the least model for each case and discuss whether it represents the participants' conclusions shown in the introduction.

4.1 Valid Arguments

$\mathcal{P}_{dog}$ represents S_{dog}. Its weak completion, $\mathsf{wc}\,\mathsf{g}\,\mathcal{P}_{dog}$, is:

$$\begin{aligned} \mathit{police\_dog}'(a) &\leftrightarrow \mathit{vicious}(a) \wedge \neg ab_{dog'}(a), \\ \mathit{police\_dog}(a) &\leftrightarrow \neg \mathit{police\_dog}'(a), \\ ab_{dog'}(a) &\leftrightarrow \bot, \\ \mathit{highly\_trained}(a) &\leftrightarrow \top, \\ \mathit{vicious}(a) &\leftrightarrow \top. \end{aligned}$$

Its least model is:

$$\langle \{\mathit{highly\_trained}(a), \mathit{vicious}(a), \mathit{police\_dog}'(a)\}, \{\mathit{police\_dog}(a), ab_{dog'}(a)\}\rangle.$$

This model entails the CONCLUSION of S_{dog}, *some highly trained dogs are not police dogs*. According to [10], S_{dog} is logically valid and psychologically believable. No conflict arises neither at the psychological nor at the logical level, and the majority concludes that this syllogism holds, which complies with the least model of $\mathsf{wc}\,\mathsf{g}\,\mathcal{P}_{dog}$.

The psychological results of the second syllogism, S_{vit}, indicate that there seems to be two kinds of participants each taking a different interpretation of the statements. The group, which validated the syllogism, was not influenced by the bias with respect to nutritional things. Accordingly, the logic program that represents their view, corresponds to $\mathcal{P}_{vit} \setminus \{ab(X) \leftarrow \mathit{vitamin}(X)\}$. The weak completion of $\mathsf{g}\,\mathcal{P}_{vit} \setminus \{ab(a) \leftarrow \mathit{vitamin}(a)\}$ is:

$$\begin{aligned} &\mathit{nutritional}'(a) \leftrightarrow \mathit{inex}(a) \wedge \neg ab(a), \\ &\mathit{nutritional}(a) \leftrightarrow \neg \mathit{nutritional}'(a), \\ &ab(a) \leftrightarrow \bot, \\ &\mathit{vitamin}(a) \leftrightarrow \top, \\ &\mathit{inex}(a) \leftrightarrow \top. \end{aligned}$$

The corresponding least model is:

$$\langle \{\mathit{vitamin}(a), \mathit{inex}(a), \mathit{nutritional}'(a)\}, \{\mathit{nutritional}(a), ab(a)\}\rangle,$$

which entails the conclusion, that *some vitamin tables are not nutritional*, and indeed we can conclude that this syllogism is valid.

The other interpretation, where participants' chose not to validate the syllogism, is the group who has apparently been influenced by their belief. Their interpretation of S_{vit} is represented by $\mathcal{P}_{vit}$. Its weak completion, $\mathsf{wc}\,\mathsf{g}\,\mathcal{P}_{vit}$, is:

$$\begin{aligned} &\mathit{nutritional}'(a) \leftrightarrow \mathit{inex}(a) \wedge \neg ab(a), \\ &\mathit{nutritional}(a) \leftrightarrow \neg \mathit{nutritional}'(a), \\ &ab(a) \leftrightarrow \bot \vee \mathit{vitamin}(a), \\ &\mathit{vitamin}(a) \leftrightarrow \top, \\ &\mathit{inex}(a) \leftrightarrow \top. \end{aligned}$$

Its least model is:

$$\langle\{\mathit{vitamin}(a), \mathit{inex}(a), \mathit{nutritional}(a), \mathit{ab}(a)\}, \{\mathit{nutritional}'(a)\}\rangle.$$

The CONCLUSION of S_{vit} is not entailed. According to [10], S_{vit} is logically valid but psychologically unbelievable. There arises a conflict at the psychological level because we generally assume that the purpose of vitamin tablets is to aid nutrition. The participants who have been influenced by this belief concluded that the syllogism does not hold, which complies with the least model of $\mathsf{lm_L wc\, g}\,\mathcal{P}_{vit}$.

4.2 Invalid Arguments

The third and the fourth cases of the syllogistic reasoning task cannot be modeled straightforwardly as the first two cases. We assume that the belief has an influence on the procedural part, that is, the reasoning process is biased. We can model this by abduction, which has been explained in Section 2.5.

$\mathcal{P}_{rich}$ represents S_{rich}. Its weak completion, $\mathsf{wc\, g}\,\mathcal{P}_{rich}$, is:

$$\begin{aligned} \mathit{mil}'(a) &\leftrightarrow \mathit{hard\_worker}(a) \wedge \neg \mathit{ab}(a), \\ \mathit{mil}(a) &\leftrightarrow \neg \mathit{mil}'(a), \\ \mathit{ab}(a) &\leftrightarrow \bot, \\ \mathit{rich}(a) &\leftrightarrow \top, \\ \mathit{hard\_worker}(a) &\leftrightarrow \top. \end{aligned}$$

Its least model is:

$$\langle\{\mathit{hard\_worker}(a), \mathit{rich}(a), \mathit{mil}'(a)\}, \{\mathit{ab}(a), \mathit{mil}(a)\}\rangle,$$

and states nothing about the CONCLUSION, *some millionaires are not rich people*. Actually, the CONCLUSION in S_{rich} states something, which contradicts PREMISE 2, and thus needs to be about something that cannot be the previously introduced constant a. According to our background knowledge, we know that millionaires exist. Let us formulate this as an observation, let's say about b: $\mathcal{O} = \{\mathit{mil}(b)\}$. If we want to allow to suppose truth or falsity of something about b with respect to $\mathcal{P}_{rich}$, say about the truth of $\mathit{hard\_worker}(b)$, we can no longer assume that $\mathsf{CONSTANTS} = \mathsf{constants}(\mathcal{P}_{rich})$, because $\mathcal{A}_{\mathsf{g}\,\mathcal{P}_{rich}}$ would not contain any facts about b. Therefore, we specify that the new set of constants in consideration is $\mathsf{CONSTANTS} = \{a, b\}$. $\mathsf{g}\,\mathcal{P}_{rich}$ with respect to $\mathsf{CONSTANTS}$ contains additionally three more clauses:

$$\begin{aligned} \mathit{mil}'(b) &\leftarrow \mathit{hard\_worker}(b) \wedge \neg \mathit{ab}(b), \\ \mathit{mil}(b) &\leftarrow \neg \mathit{mil}'(b), \\ \mathit{ab}(b) &\leftarrow \bot. \end{aligned}$$

The set of abducibles, $\mathcal{A}_{\mathsf{g}\,\mathcal{P}_{rich}}$, contains the following clauses:

$$\mathit{hard\_worker}(b) \leftarrow \top, \qquad \mathit{hard\_worker}(b) \leftarrow \bot.$$

$\mathcal{E} = \{hard\_worker(b) \leftarrow \bot\}$ is the only explanation for $\mathcal{O}$. $\mathsf{wc}\,\mathsf{g}\,(\mathcal{P}_{rich} \cup \mathcal{E})$ contains:

$$\begin{aligned} mil'(b) &\leftrightarrow hard\_worker(b) \wedge \neg ab(b), \\ mil(b) &\leftrightarrow \neg mil'(b), \\ ab(b) &\leftrightarrow \bot, \\ hard\_worker(b) &\leftrightarrow \bot. \end{aligned}$$

Its least model, where $\mathsf{lm}_{\text{L}}\mathsf{wc}\,\mathsf{g}\,(\mathcal{P}_{rich} \cup \mathcal{E}) = \langle I^\top, I^\bot \rangle$, contains:

$$\begin{aligned} I^\top &= \{mil(b)\}, \\ I^\bot &= \{ab(b), mil'(b), hard\_worker(b)\}. \end{aligned}$$

As this model does not confirm the CONCLUSION it does not validate S_{rich}. According to [10] this case is quite easy to solve, because it is neither logically valid nor believable. Almost no one validated S_{rich}, which complies with the least model of $\mathsf{wc}\,\mathsf{g}\,(\mathcal{P}_{rich} \cup \mathcal{E})$.

$\mathcal{P}_{cig}$ represents S_{cig}. Its weak completion, $\mathsf{wc}\,\mathsf{g}\,\mathcal{P}_{cig}$, is:

$$\begin{aligned} addictive'(a) &\leftrightarrow inex(a) \wedge \neg ab_{add'}(a), \\ addictive(a) &\leftrightarrow \neg addictive'(a), \\ ab_{add'}(a) &\leftrightarrow \bot \vee cig(a), \\ cig(a) &\leftrightarrow \top, \\ inex(a) &\leftrightarrow (cig(a) \wedge \neg ab_{inex}(a)) \vee \top, \\ ab_{inex}(a) &\leftrightarrow \bot. \end{aligned}$$

Its least model of the weak completion is:

$$\langle \{cig(a), inex(a), addictive(a), ab_{add'}(a)\}, \{addictive'(a), ab_{inex}(a)\} \rangle,$$

which, similarly to the previous case, does not state anything about the CONCLUSION, *some addictive things are not cigarettes*. Again, the CONCLUSION of S_{cig} is something, which cannot be about a. According to our background knowledge, we know that addictive things exist. Let us formulate this again as an observation, say about b: $\mathcal{O} = \{addictive(b)\}$, which needs to be explained. In order to generate an explanation for $\mathcal{O}$, let us define $\mathsf{CONSTANTS} = \{a, b\}$. $\mathsf{g}\,\mathcal{P}_{rich}$ with respect to $\mathsf{CONSTANTS}$ now additionally contains five more clauses:

$$\begin{aligned} addictive'(b) &\leftarrow inex(b) \wedge \neg ab_{add'}(b), \\ addictive(b) &\leftarrow \neg addictive'(b), \\ ab_{add'}(b) &\leftarrow \bot, \\ ab_{add'}(b) &\leftarrow cig(b), \\ inex(b) &\leftarrow cig(b) \wedge \neg ab_{inex}(b), \\ ab_{inex}(b) &\leftarrow \bot. \end{aligned}$$

Given $\mathsf{g}\,\mathcal{P}_{cig}$, the set of abducibles, $\mathcal{A}_{\mathsf{g}\,\mathcal{P}_{cig}}$, contains the following clauses:

$$cig(b) \leftarrow \top, \qquad cig(b) \leftarrow \bot.$$

$\mathcal{O}$ is true if $addictive'(b)$ is false, which is false if $inex(b)$ is false or $ab_{add'}(b)$ is true. $inex(b)$ is false if $cig(b)$ is false and $ab_{add'}(b)$ is true if $cig(b)$ is true. For $\mathcal{O}$ we have two minimal explanations, $\mathcal{E}_\bot = \{cig(b) \leftarrow \bot\}$ and $\mathcal{E}_\top = \{cig(b) \leftarrow \top\}$. The weak completion of $\mathsf{g}\,(\mathcal{P}_{cig} \cup \mathcal{E}_\bot)$ contains:

$$
\begin{aligned}
addictive'(b) &\leftrightarrow inex(b) \wedge \neg ab_{add'}(b),\\
addictive(b) &\leftrightarrow \neg addictive'(b),\\
ab_{add'}(b) &\leftrightarrow \bot \vee cig(b),\\
inex(b) &\leftrightarrow cig(b) \wedge \neg ab_{inex}(b),\\
ab_{inex}(b) &\leftrightarrow \bot,\\
cig(b) &\leftrightarrow \bot.
\end{aligned}
$$

Its least model, where $\mathsf{lm}_{\text{L}}\mathsf{wc}\,\mathsf{g}\,(\mathcal{P}_{cig} \cup \mathcal{E}_\bot) = \langle I^\top, I^\bot \rangle$ contains:

$$
\begin{aligned}
I^\top &= \{addictive(b)\},\\
I^\bot &= \{cig(b), inex(b), ab_{add'}(b), ab_{inex}(b)\},
\end{aligned}
$$

which entails the CONCLUSION of S_{cig}. As $\mathcal{E}_\top$ is yet another explanation for $\mathcal{O}$, the CONCLUSION, that b is not a cigarette, only follows credulously. S_{cig} is logically invalid but psychologically believable and therefore causes a conflict [10]: S_{cig} does not follow logically from the premises; however, people are biased and search for a model, which confirms their beliefs. Therefore, the majority concluded that this syllogism holds, which complies with the least model of $\mathsf{wc}\,\mathsf{g}\,(\mathcal{P}_{add} \cup \mathcal{E}_\bot)$.
In [26,27], we show an extension of this case, where the conclusion follows skeptically. With help of meta predicates, we specify that the first premise describes the usual and the second premise describes the exceptional case. That is, an inexpensive cigarette is meant to be the exception not the rule, in the context of things that are addictive and expensive.

5 Conclusion

The weak completion semantics has shown to successfully model various human reasoning episodes [4,5,7,18,26,27]. This paper presents yet another human reasoning task modeled under the weak completion semantics. As in our previous formalizations, we follow Stenning and van Lambalgen's two step approach. We motivate our assumptions based on results from Psychology, where syllogisms in human reasoning have been investigated extensively in the past decades.
As has been shown in the previous formalizations, the advantage of the weak completion semantics over other logic programming approaches, is, that undefined atoms stay unknown, instead of becoming false. The syllogistic reasoning tasks, which have been discussed in the literature so far, have never accounted to give the option 'I don't know' to the participants. As has been discussed in [24], participants who say that no valid conclusion follows, might have problems to actually find a conclusion easily and possibly mean that they simply do not know.

They also point to [28], who suggest that, if a conclusion is stated as being not valid, this could just simply mean that the reasoning process is exhausted. An experimental study, which would allow the participants to distinguish between 'I don't know' and 'not valid', might possibly give us more insights about their reasoning processes and identify where exactly the belief bias takes effect.

6 Acknowledgements

Many thanks to Steffen Hölldobler and Luís Moniz Pereira for valuable feedback.

References

1. J. Adler and L. Rips. *Reasoning: Studies of Human Inference and Its Foundations.* Cambridge University Press, 2008.
2. L. J. Chapman and J. P. Chapman. Atmosphere effect re-examined. *Journal of Experimental Psychology*, 58(3):220–6, 1959.
3. K. L. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, volume 1, pages 293–322. Plenum Press, New York, NY, 1978.
4. E.-A. Dietz, S. Hölldobler, and M. Ragni. A computational logic approach to the suppression task. In N. Miyake, D. Peebles, and R. P. Cooper, editors, *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pages 1500–1505, Austin, TX, 2012.
5. E.-A. Dietz, S. Hölldobler, and M. Ragni. A computational logic approach to the abstract and the social case of the selection task. In *11th International Symposium on Logical Formalizations of Commonsense Reasoning*, 2013.
6. E.-A. Dietz, S. Hölldobler, and C. Wernhard. Modeling the suppression task under weak completion and well-founded semantics. *Journal of Applied Non-Classical Logics*, 2013.
7. E.-A. Dietz, S. Hölldobler, and C. Wernhard. Modeling the suppression task under weak completion and well-founded semantics. *Journal of Applied Non-Classsical Logics*, 24(1–2):61–85, 2014.
8. J. Evans. Thinking and believing. *Mental models in reasoning*, 2000.
9. J. Evans. Biases in deductive reasoning. In R. Pohl, editor, *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory.* Psychology Press, 2012.
10. J. Evans, J. L. Barston, and P. Pollard. On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3):295–306, 1983.
11. J. Evans, S. Handley, and C. Harper. Necessity, possibility and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, 54(3):935–958, 2001.
12. J. S. Evans. *Bias in human reasoning - causes and consequences.* Essays in cognitive psychology. Lawrence Erlbaum, 1989.
13. M. Fitting. A Kripke-Kleene semantics for logic programs. *Journal of Logic Programming*, 2(4):295–312, 1985.
14. A. Garnham and J. Oakhill. *Thinking and Reasoning.* Wiley, 1994.
15. S. Hölldobler. *Logik und Logikprogrammierung 1: Grundlagen.* Kolleg Synchron. Synchron, 2009.

16. S. Hölldobler and C. D. Kencana Ramli. Logic programs under three-valued Łukasiewicz semantics. In P. M. Hill and D. S. Warren, editors, *Logic Programming, 25th International Conference, ICLP 2009*, volume 5649 of *Lecture Notes in Computer Science*, pages 464–478, Heidelberg, 2009. Springer.
17. S. Hölldobler and C. D. Kencana Ramli. Logics and networks for human reasoning. In C. Alippi, M. M. Polycarpou, C. G. Panayiotou, and G. Ellinas, editors, *International Conference on Artificial Neural Networks, ICANN 2009, Part II*, volume 5769 of *Lecture Notes in Computer Science*, pages 85–94, Heidelberg, 2009. Springer.
18. S. Hölldobler, T. Philipp, and C. Wernhard. An abductive model for human reasoning. In *Logical Formalizations of Commonsense Reasoning, Papers from the AAAI 2011 Spring Symposium*, AAAI Spring Symposium Series Technical Reports, pages 135–138, Cambridge, MA, 2011. AAAI Press.
19. P. N. Johnson-Laird. *Mental models: towards a cognitive science of language, inference, and consciousness*. Harvard University Press, Cambridge, MA, 1983.
20. P. N. Johnson-Laird and R. M. Byrne. Deduction. 1991.
21. A. C. Kakas, R. A. Kowalski, and F. Toni. Abductive logic programming. *Journal of Logic and Computation*, 2(6):719–770, 1993.
22. J. W. Lloyd. *Foundations of Logic Programming*. Springer-Verlag New York, Inc., New York, NY, USA, 1984.
23. J. Łukasiewicz. O logice trójwartościowej. *Ruch Filozoficzny*, 5:169–171, 1920. English translation: On three-valued logic. In: Łukasiewicz J. and Borkowski L. (ed.). (1990). *Selected Works*, Amsterdam: North Holland, pp. 87–88.
24. S. Newstead, S. Handley, and E. Buck. Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory & Cognition*, 27(2):344–354, 1999.
25. S. E. Newstead and R. A. Griggs. Fuzzy quantifiers as an explanation of set inclusion performance. *Psychological Research*, 46(4):377–388, 1984.
26. L. M. Pereira, E.-A. Dietz, and S. Hölldobler. A computational logic approach to the belief bias effect. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning*, 2014.
27. L. M. Pereira, E.-A. Dietz, and S. Hölldobler. Contextual abductive reasoning with side-effects. volume 14, pages 633–648, 2014.
28. T. A. Polk and A. Newell. Deduction as verbal reasoning. *Psychological Review*, 102(3):533–566, 1995.
29. K. Stenning and M. van Lambalgen. Semantic interpretation as computation in nonmonotonic logic: The real meaning of the suppression task. *Cognitive Science*, 6(29):916–960, 2005.
30. K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science*. A Bradford Book. MIT Press, Cambridge, MA, 2008.
31. A. Van Gelder, K. A. Ross, and J. S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):619–649, 1991.
32. M. Wilkins. he effect of changed material on the ability to do formal syllogistic reasoning. 16(102), 1928.
33. R. S. Woodworth and S. B. Sells. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 18(4):451–60, 1935.