A Computational Logic Approach to Human Syllogistic Reasoning

Ana Oliveira da Costa* and Emmanuelle-Anna Dietz Saldanha* and Steffen Hölldobler* and Marco Ragni**[1]
* International Center for Computational Logic, TU Dresden, Germany, ana.oli.costa@gmail.com, {dietz,sh}@iccl.tu-dresden.de
** Cognitive Computation lab, University of Freiburg, Germany, ragni@cs.uni-freiburg.de

Abstract

A recent meta-analysis (Khemlani & Johnson-Laird, 2012) about psychological experiments of syllogistic reasoning demonstrates that the conclusions drawn by human reasoners strongly deviate from conclusions of classical logic. Moreover, none of the current cognitive theories predictions fit reliably the empirical data. In this paper, we show how human syllogistic reasoning can be modeled under a new cognitive theory, the Weak Completion Semantics. Our analysis based on computational logics identifies seven principles necessary to draw the inferences. Hence, this work contributes to a computational foundation of cognitive reasoning processes.

Keywords: Human Reasoning; Syllogistic Reasoning; Logic Programming; Three-valued Łukasiewicz Logic; Abduction

Introduction

The way of how humans ought to reason correctly about syllogisms has already been investigated by Aristotle (Smith, 1989). A syllogism consists of two premises and a conclusion. Each of them is a quantified statement using one of the four quantifiers *all* (A), *no* (E), *some* (I), and *some are not* (O)[2] about sets of entities which we denote in the following by *a*, *b*, and *c*. An example of a syllogism is:

Some b are a (IE4)
No b are c

In experiments, participants are normally expected to complete the syllogism by drawing a logical consequence from the first two premises, e.g. in this example '*some a are not c*'. The response of the participant – the conclusion – is evaluated as true if it can be derived in classical first-order logic (FOL), otherwise as false. The four quantifiers and their formalization in FOL are given in Table 1. The entities can appear in four different orders called *figures* as shown in Table 2. Hence, a problem can be completely specified by the quantifiers of the first and second premise and the figure. The example discussed above is denoted by IE4.

Altogether, there are 64 syllogisms and, if formalized in FOL, we can compute their logical consequence in classical logic. However, a meta-analysis by Khemlani and Johnson-Laird (2012) based on six experiments has shown that humans do not only systematically deviate from the predictions of FOL but from any other of 12 cognitive theories. In the case of IE4, besides the above mentioned logical consequence, a significant number of humans answered '*no valid conclusion*', which does not follow from IE4 in FOL, as *'some a are not c'* follows from IE4 as can be seen in the Venn diagram in Figure 1: *X* has the property *a* but not the property *c*.

[1]The authors are mentioned in alphabetical order.
[2]We are using the classical abbreviations.

Table 1: The four moods and their logical formalization.

Mood	Natural language	First-order logic	Short
affirmative universal	*all a are b*	$\forall X(a(X) \rightarrow b(X))$	Aab
affirmative existential	*some a are b*	$\exists X(a(X) \wedge b(X))$	Iab
negative universal	*no a are b*	$\forall X(a(X) \rightarrow \neg b(X))$	Eab
negative existential	*some a are not b*	$\exists X(a(X) \wedge \neg b(X))$	Oab

Table 2: The four figures used in syllogistic reasoning.

	Figure 1	Figure 2	Figure 3	Figure 4
First Premise	a-b	b-a	a-b	b-a
Second Premise	b-c	c-b	c-b	b-c

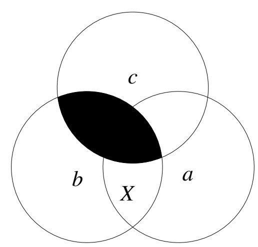

Figure 1: '*some a are not c*' follows from IE4.

Recently, a new cognitive theory based on the Weak Completion Semantics (WCS) has been developed (Hölldobler, 2015). It has its roots in the ideas first expressed by Stenning and van Lambalgen (2008), which unfortunately had some technical mistakes. These were corrected by Hölldobler and Kencana Ramli (2009a), by using the three-valued Łukasiewicz (1920) logic. WCS has been successfully applied – among others – to the suppression task (Dietz, Hölldobler, & Ragni, 2012), the selection task (Dietz, Hölldobler, & Ragni, 2013), the belief-bias effect (Pereira, Dietz, & Hölldobler, 2014a, 2014b; Dietz, 2017), to reasoning about conditionals (Dietz & Hölldobler, 2015; Dietz, Hölldobler, & Pereira, 2015) and to spatial reasoning (Dietz, Hölldobler, & Höps, 2015). Hence, it was natural to ask whether WCS can also model syllogistic reasoning.

We introduce seven principles motivated by findings made in Cognitive Science and Computational Logic, and show how syllogisms together with these principles can be encoded in logic programs. After that we compare the predictions under WCS with the results of FOL, the syntactic rule based theory PSYCOP (Rips, 1994), the Verbal Model Theory (Polk & Newell, 1995) and the Mental Model Theory (Johnson-Laird, 1983).[3] The two model-based theories performed best in the meta-analysis (Khemlani & Johnson-Laird, 2012).

The predictions of the theories FOL, PSYCOP, Verbal, and Mental Models for the syllogisms OA4, IE4, and IA2 and those of a significant percentage of the participants are depicted in Table 3, where the participants were 156

[3]http://mentalmodels.princeton.edu/models/mreasoner/

Table 3: Conclusions drawn by a significant percentage of participants are highlighted in gray and compared to the predictions of FOL, PSYCOP, Verbal, and Mental Models theory for OA4, IE4 and IA2. NVC denotes *no valid conclusion*.

	Participants	FOL	PSYCOP	Verbal Models	Mental Models
OA4	Oca	Oca	Oca, Ica, Iac	Oca, NVC	Oca, Oac, NVC
IE4	Oac, NVC	Oac	Oac, Iac, Ica	NVC, Oac	Oac, NVC Eac, Eca, Oca
IA2	Iac, Ica	NVC	NVC	NVC, Ica	Iac, Ica, NVC

high school or university students. A conclusion is drawn by a significant number of participants, if the number of participants chosing this particular conclusion is statistically too high for the conclusion to be chosen at random. The interested reader is referred to Khemlani and Johnson-Laird (2012) for more details.

FOL and the other three cognitive theories make different predictions. Each theory provides at least one prediction which is correct with respect to FOL and provides an additional prediction which is false with respect to FOL. Currently, the best results are achieved by the Verbal Models Theory which predicts 84% of the participants responses, closely followed by the Mental Model Theory with 83%, whereas PSYCOP predicts 77% of the participants responses.

Weak Completion Semantics

Logic Programs

A *(logic) program* $\mathcal{P}$ is a finite set of clauses of the form $A \leftarrow \top$, $A \leftarrow \bot$ or $A \leftarrow B_1 \wedge \ldots \wedge B_n$, $n > 0$, where A is an atom, B_i, $1 \leq i \leq n$, are literals, and $\top$ and $\bot$ denote truth and falsehood, respectively. Clauses are assumed to be universally closed. A is called *head* and $\top$, $\bot$ as well as $B_1 \wedge \ldots \wedge B_n$ are called *body* of the corresponding clause. Clauses of the form $A \leftarrow \top$ and $A \leftarrow \bot$ are called *facts* and *assumptions*, respectively.[4] $\neg A$ is *assumed* in $\mathcal{P}$ iff $\mathcal{P}$ contains an assumption with head A and no other clause with head A occurs in $\mathcal{P}$. We restrict terms to be constants and variables. We assume for each program $\mathcal{P}$ that the underlying alphabet consists precisely of the symbols occurring in $\mathcal{P}$ and that non-propositional programs contain at least one constant.

$g\mathcal{P}$ denotes the set of all ground instances of clauses occurring in $\mathcal{P}$, where a ground instance of clause C is obtained from C by replacing each variable occurring in C by a constant. A ground atom A is *defined* in $g\mathcal{P}$ iff $g\mathcal{P}$ contains a clause whose head is A; otherwise A is said to be *undefined*. $def(A,\mathcal{P}) = \{A \leftarrow Body \mid A \leftarrow Body \in g\mathcal{P}\}$ is called *definition* of A in $\mathcal{P}$. To clarify the definitions, consider $\mathcal{P}_{ex}$:

$$p(X) \leftarrow q(X) \wedge \neg r(X) \wedge s(X). \qquad q(a) \leftarrow \top. \qquad r(a) \leftarrow \bot.$$

[4] $A \leftarrow \bot$ is called an assumption because it can be overwritten under the Weak Completion Semantics, as we will discuss later.

Table 4: The truth tables for the connectives under three-valued Łukasiewicz logic. Note that for the unknown holds: $(U \leftarrow U) = \top$.

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

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

$\leftarrow$	$\top$	U	$\bot$
$\top$	$\top$	$\top$	$\top$
U	U	$\top$	$\top$
$\bot$	$\bot$	U	$\top$

The second and the third clause is a fact and an assumption, respectively. $g\mathcal{P}_{ex}$ is as follows:

$$p(a) \leftarrow q(a) \wedge \neg r(a) \wedge s(a). \qquad q(a) \leftarrow \top. \qquad r(a) \leftarrow \bot.$$

$p(a), q(a)$, $r(a)$ are defined and $s(a)$ is undefined in $g\mathcal{P}_{ex}$. Classical logic and logic programs are discussed in detail in e.g. (Hölldober, 2009; Lloyd, 1984).

Three-Valued Łukasiewicz Logic

We consider the three-valued Łukasiewicz (1920) logic for which the corresponding truth values are *true* ($\top$), *false* ($\bot$) and *unknown* (U). A *three-valued interpretation* I is a mapping from the set of formulas to the set $\{\top, \bot, U\}$. The truth value of a given formula under I is determined according to the truth tables in Table 4. We represent an interpretation as a pair $I = \langle I^\top, I^\bot \rangle$ of disjoint sets of ground 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 formula F is mapped to true under I. $\mathcal{M}$ is a *model* of $\mathcal{P}$ if it is an interpretation, which maps each clause occurring in $g\mathcal{P}$ to $\top$. I is the *least model* of $\mathcal{P}$ iff for any other model J of $\mathcal{P}$ it holds that $I \subseteq J$.

Least Models under the Weak Completion

For a given program $\mathcal{P}$, consider the following transformation: (1) For each ground atom A which is defined in $g\mathcal{P}$, replace all clauses of the form $A \leftarrow Body_1, \ldots, A \leftarrow Body_m$ occurring in $g\mathcal{P}$ by $A \leftarrow Body_1 \vee \ldots \vee Body_m$. (2) Replace all occurrences of $\leftarrow$ by $\leftrightarrow$. The obtained set of formulas is called *weak completion* of $\mathcal{P}$ or $wc\mathcal{P}$.[5]

It has been shown by Hölldobler and Kencana Ramli (2009b) that programs as well as their weak completions admit a least model under three-valued Łukasiewicz logic. Moreover, the least model of $wc\mathcal{P}$ can be obtained as the least fixed point of the semantic operator Φ, which is due to Stenning and van Lambalgen (2008). Let $I = \langle I^\top, I^\bot \rangle$ be an interpretation, then $\Phi_{\mathcal{P}}(I) = \langle J^\top, J^\bot \rangle$, is defined by:

$$\begin{array}{rcl} J^\top & = & \{A \mid A \leftarrow Body \in def(A,\mathcal{P}) \text{ and } I(Body) = \top\}, \\ J^\bot & = & \{A \mid def(A,\mathcal{P}) \neq \emptyset \text{ and} \\ & & \quad I(Body) = \bot \text{ for all } A \leftarrow Body \in def(A,\mathcal{P})\}. \end{array}$$

Weak Completion Semantics (WCS) *is the approach to consider weakly completed programs, to compute their least*

[5] If $\mathcal{P} = \{A \leftarrow \bot, A \leftarrow \top\}$ then $wc\mathcal{P} = \{A \leftrightarrow \bot \vee \top\}$. This is semantically equivalent to $wc\mathcal{P} = \{A \leftrightarrow \top\}$. $A \leftarrow \bot$ is overwritten.

models, and to reason with respect to these models.

We write $\mathcal{P} \models_{wcs} F$ iff formula F holds in the least model of $wc\mathcal{P}$. Consider again $\mathcal{P}_{\text{ex}}$. Starting with $\langle\emptyset,\emptyset\rangle$, we obtain:

$$\Phi_{\mathcal{P}_{\text{ex}}}(\langle\emptyset,\emptyset\rangle) \quad = \quad \langle\{q(a)\},\{r(a)\}\rangle \quad = \quad I_1 \quad = \quad \Phi_{\mathcal{P}_{\text{ex}}}(I_1).$$

I_1 is the least model of $wc\mathcal{P}_{\text{ex}}$.

Integrity Constraints

An *integrity constraint* is an expression of the form $\mathsf{U} \leftarrow Body$, where *Body* is a conjunction of literals and U denotes the unknown. An interpretation I maps an integrity constraint $\mathsf{U} \leftarrow Body$ to $\top$ iff $I(Body) \subseteq \{\bot,\mathsf{U}\}$. Given an interpretation I and a finite set IC of integrity constraints, I *satisfies* IC iff all clauses occurring in IC are true under I.

Abductive Logic Programming

In Logic Programming, *abduction* is the reasoning process of searching for explanations given a program and some observations, which do not follow from the program (Peirce, Hartshorne, & Weiss, 1974). Explanations are usually restricted to certain formulas called *abducibles*. Let $\mathcal{P}$ be a ground program. The *set of abducibles of* $\mathcal{P}$ is

$$\begin{array}{rcl} \mathcal{A}_{\mathcal{P}} & = & \{A \leftarrow \top \mid A \text{ is undefined or } \neg A \text{ is assumed in } \mathcal{P}\} \\ & \cup & \{A \leftarrow \bot \mid A \text{ is undefined in } \mathcal{P}\}. \end{array}$$

An *abductive framework* consists of a program $\mathcal{P}$, a finite set $\mathcal{A}$ of abducibles, a finite set IC of integrity constraints, and an entailment relation. Let $\langle\mathcal{P},\mathcal{A},\mathsf{IC},\models_{wcs}\rangle$ be an abductive framework, $\mathcal{E} \subseteq \mathcal{A}$, and O a non-empty set of literals called *observation*. An observation $O = \{o_1,\ldots,o_n\}$ is *explained by* $\mathcal{E}$ *given* $\mathcal{P}$ *and* IC iff $\mathcal{P} \cup \mathcal{E} \models_{wcs} o_1 \wedge \ldots \wedge o_n$ and $\mathcal{P} \cup \mathcal{E} \models_{wcs} \mathsf{IC}$. O is *explained given* $\mathcal{P}$ *and* IC iff there exists an $\mathcal{E}$ such that O is explained by $\mathcal{E}$ given $\mathcal{P}$ and IC. We prefer subset-minimal explanations. An explanation $\mathcal{E}$ is minimal iff there is no explanation $\mathcal{E}'$ such that $\mathcal{E}' \subset \mathcal{E}$.

Consider the program $\mathcal{P} = \{w \leftarrow g,\ w \leftarrow s,\ g \leftarrow r\}$.[6] Suppose we observe $O = \{w\}$. Because the least model of $\mathsf{wc}\mathcal{P}$ is $\langle\emptyset,\emptyset\rangle$ the observation does not follow. However, O can be explained by $\mathcal{E} = \{s \leftarrow \top\}$. Starting with the empty interpretation we obtain:

$$\begin{array}{rcl} \Phi_{\mathcal{P}\cup\mathcal{E}}(\langle\emptyset,\emptyset\rangle) & = & \langle\{s\},\emptyset\rangle \\ \Phi_{\mathcal{P}\cup\mathcal{E}}(\langle\{s\},\emptyset\rangle) & = & \langle\{s,w\},\emptyset\rangle \quad = \quad \Phi_{\mathcal{P}\cup\mathcal{E}}(\langle\{s,w\},\emptyset\rangle). \end{array}$$

$\langle\{s,w\},\emptyset\rangle$ is the least model of $wc(\mathcal{P}\cup\mathcal{E})$. It entails O. $\mathcal{E}$ is minimal, whereas $\mathcal{E}' = \{s \leftarrow \top, r \leftarrow \top\}$ is not.

[6] The wheels of a lawnmower are wet if the gras is wet; the wheels are wet if the sprinkler is on; the gras is wet if it is raining.

Seven Principles in Reasoning with Quantifiers

We will apply seven principles in developing a logical form for the representation of syllogisms.

Licenses for Inferences (licenses)

Stenning and van Lambalgen (2008) propose to formalize conditionals by *licenses for inferences*. For example, the conditional *for all X, if* $p(X)$ *then* $q(X)$ is represented by the program $\{q(X) \leftarrow p(X) \wedge \neg ab(X),\ ab(X) \leftarrow \bot\}$. Its first clause states that for all X, $q(X)$ holds if $p(X)$ and $\neg ab(X)$ holds, i.e. nothing abnormal for X is known. This in turn is stated by the second clause. Clauses are assumed to be universally closed and, hence, the universal quantifier can be omitted.

Existential Import and Gricean Implicature (import)

Humans seem to understand quantifiers differently due to a pragmatic understanding of language. For instance, in natural language we normally do not quantify over things that do not exist. Consequently, *for all* implies *there exists*. This appears to be in line with human reasoning and has been called the *Gricean Implicature* (Grice, 1975). Several theories like the theory of mental models (Johnson-Laird, 1983) or mental logic (Rips, 1994) assume that the sets we quantify over are not empty. Likewise, Stenning and van Lambalgen (2008) have shown that humans require existential import for a conditional to be true. Furthermore, as mentioned in Khemlani and Johnson-Laird (2012), the quantifier '*some a are b*' often implies that '*some a are not b*', which again can be explained by assuming the *Gricean Implicature*: Someone would not state '*some a are b*' if that person knew that '*all a are b*'. As the person does not say '*all a are b*' but instead '*some a are b*', we have to assume that '*not all a are b*', which in turn implies '*some a are not b*'.

Unknown Generalization (unknownGen)

Humans seem to distinguish between '*some y are z*' and '*all y are z*', as we have already explained in the previous paragraph. Accordingly, if we observe that an object o belongs to y and z then we do not want to conclude both, '*some y are z*' and '*all y are z*'. In order to prevent such unwanted conclusions we introduce the following principle: if we know that '*some y are z*' then there must not only be an object o_1 which belongs to y and z (by Gricean implicature) but there must be another object o_2 which belongs to y and for which it is unknown whether it belongs to z.

Converse Interpretation (converse)

Although there appears to be some evidence that humans seem to distinguish between '*some y are z*' and '*some z are y*' (see the results reported in Khemlani & Johnson-Laird, 2012) we propose that premises of the form Iab imply Iba and vice versa. If there is an object which belongs to y and z, then there is also an object which belongs to z and y.

Block Conclusions by Double Negation (doubleNeg)

Consider the following two negative sentences (i.e. including negation) and the positive one: ' If not a, then b. If not b then c. a is true.' The program representing these sentences is $\mathcal{P} = \{b \leftarrow \neg a,\ c \leftarrow \neg b,\ a \leftarrow \top\}$. The weak completion of $\mathcal{P}$ is $\mathsf{wc}\mathcal{P} = \{b \leftrightarrow \neg a,\ c \leftrightarrow \neg b,\ a \leftrightarrow \top\}$. Its least model is $\langle\{a,c\},\{b\}\rangle$, and thus a and c are true: a is true because it is a fact and c is true by the negation of b. b is derived to be false because the negation of a is false. This example shows that under WCS, a positive conclusion (c being true) can be derived from two clauses, which include negation. Considering the results of the participants' responses in (Khemlani & Johnson-Laird, 2012), they seem not to draw conclusions through double negation. Accordingly, we block them through abnormalities.

Search Alternative Conclusions to NVC (abduction)

Our hypothesis is that when participants are faced with a NVC conclusion (*'no valid conclusion'*), they might not want to accept this conclusion and proceed to check whether there exists unknown information that is relevant. This information may be explanations to facts in our program, and we model such repair mechanism using abduction. Facts in our programs come either from an existential import or from unknown generalization. We use only the first as source for observations, since they are used directly to infer new information.

Negation by Transformation (transformation)

A negative literal cannot be the head of a clause in a program. In order to represent a negative conclusion $\neg p(X)$ an auxiliary atom $p'(X)$ is introduced together with a clause $p(X) \leftarrow \neg p'(X)$ and the integrity constraint $\mathsf{U} \leftarrow p(X) \wedge p'(X)$. This is a widely used technique in logic programming. Together with the principle *licences for inferences*, the additional clause is $p(X) \leftarrow \neg p'(X) \wedge \neg ab(X)$.

Representation of Quantified Statements

Based on the principles presented in the previous section, we discuss the representation of three examples.

OA4 The two syllogistic premises of OA4 are as follows:

Some b are not a. (Oba) *all b are c. (Abc)*

The program $\mathcal{P}_{\mathsf{OA4}}$ representing OA4 consists of:

$a'(X) \leftarrow b(X) \wedge \neg ab_{bna}(X).$ (transformation &licenses)
$ab_{bna}(o_1) \leftarrow \bot.$ (unknownGen & licenses)
$a(X) \leftarrow \neg a'(X) \wedge \neg ab_{naa}(X).$ (transformation & licenses)
$b(o_1) \leftarrow \top.$ (import)
$b(o_2) \leftarrow \top.$ (unknownGen)
$ab_{naa}(o_1) \leftarrow \bot.$ (doubleNeg & licenses)
$ab_{naa}(o_2) \leftarrow \bot.$ (doubleNeg & licenses)

$c(X) \leftarrow b(X) \wedge \neg ab_{bc}(X).$ (licenses)
$ab_{bc}(X) \leftarrow \bot.$ (licenses)
$b(o_3) \leftarrow \top.$ (import)

In addition, we have the following integrity constraint:

$\mathsf{U} \leftarrow a(X) \wedge a'(X).$ (transformation)

IE4 The two syllogistic premises of IE4 are as follows:

Some b are a. (Iba) *no b are c. (Ebc)*

The program $\mathcal{P}_{\mathsf{IE4}}$ representing IE4 consists of:

$a(X) \leftarrow b(X) \wedge \neg ab_{ba}(X).$ (licenses)
$ab_{ba}(o_1) \leftarrow \bot.$ (licenses&unknownGen)
$b(o_1) \leftarrow \top.$ (import)
$b(o_2) \leftarrow \top.$ (unknownGen)

$b(X) \leftarrow a(X) \wedge \neg ab_{ab}(X).$ (converse&licenses)
$ab_{ab}(o_3) \leftarrow \bot.$ (converse&licenses&unknownGen)
$a(o_3) \leftarrow \top.$ (converse&import)
$a(o_4) \leftarrow \top.$ (converse&unknownGen)

$c'(X) \leftarrow b(X) \wedge \neg ab_{bnc}(X).$ (transformation &licenses)
$ab_{bnc}(X) \leftarrow \bot.$ (licenses)
$c(X) \leftarrow \neg c'(X) \wedge \neg ab_{ncc}(X).$ (transformation &licenses)
$b(o_5) \leftarrow \top.$ (import)
$ab_{ncc}(o_5) \leftarrow \bot.$ (licenses&doubleNeg)

In addition, we have the following integrity constraint:

$\mathsf{U} \leftarrow c(X) \wedge c'(X).$ (transformation)

IA2 The two syllogistic premises of IA2 are as follows:

Some b are a. (Iba) *all c are b. (Acb)*

Program $\mathcal{P}_{\mathsf{IA2}}$ consists of the following clauses:

$a(X) \leftarrow b(X) \wedge \neg ab_{ba}(X).$ (licenses)
$ab_{ba}(o_1) \leftarrow \bot.$ (licenses&unknownGen)
$b(o_1) \leftarrow \top.$ (import)
$b(o_2) \leftarrow \top.$ (unknownGen)

$b(X) \leftarrow a(X) \wedge \neg ab_{ab}(X).$ (converse&licenses)
$ab_{ab}(o_3) \leftarrow \bot.$ (converse&licenses&unknownGen)
$a(o_3) \leftarrow \top.$ (converse&import)
$a(o_4) \leftarrow \top.$ (converse&unknownGen)

$b(X) \leftarrow c(X) \wedge \neg ab_{cb}(X).$ (licenses)
$ab_{cb}(X) \leftarrow \bot.$ (licenses&unknownGen)
$c(o_5) \leftarrow \top.$ (import)

Entailment of Syllogisms

Khemlani and Johnson-Laird (2012) do not report a formal definition for the entailment of syllogisms. They use FOL as a normative theory and test if the conclusions drawn by the participants are correct with respect to a first-order representation of a syllogism. In FOL, a set of formulas $\mathcal{K}$ *entails* a formula F ($\mathcal{K} \models F$) if all interpretations which map $\mathcal{K}$ to true map F to true as well. We believe that the entailment relation should reflect the principles on which the representation is based. In the following, an entailment relation regarding WCS is presented, where yz is to be replaced by ac or ca.

- $\mathcal{P} \models Ayz$ iff there exists an object o such that $\mathcal{P} \models_{wcs} y(o)$ and for all o we find that if $\mathcal{P} \models_{wcs} y(o)$ then $\mathcal{P} \models_{wcs} z(o)$.
- $\mathcal{P} \models Eyz$ iff there exists an object o such that $\mathcal{P} \models_{wcs} z(o)$ and for all o we find that if $\mathcal{P} \models_{wcs} z(o)$ then $\mathcal{P} \models_{wcs} \neg y(o)$.
- $\mathcal{P} \models Iyz$ iff there exists an object o_1 such that $\mathcal{P} \models_{wcs} y(o_1) \wedge z(o_1)$ and there exists an object o_2 such that $\mathcal{P} \models_{wcs} y(o_2)$ and $\mathcal{P} \not\models_{wcs} z(o_2)$ and there exists an object o_3 such that $\mathcal{P} \models_{wcs} z(o_3)$ and $\mathcal{P} \not\models_{wcs} y(o_3)$.
- $\mathcal{P} \models Oyz$ iff there exists an object o_1 such that $\mathcal{P} \models_{wcs} y(o_1) \wedge \neg z(o_1)$ and there exists an object o_2 such that $\mathcal{P} \models_{wcs} y(o_2)$ and $\mathcal{P} \not\models_{wcs} \neg z(o_2)$.

In case we can not conclude any of these moods, then no valid conclusion is entailed, denoted as $\mathcal{P} \models$ NVC.

Accuracy of Predictions

We evaluate our approach the same way as the other theories have been evaluated (Khemlani & Johnson-Laird, 2012): There are nine different answers for each of the 64 syllogisms that can be ordered in a list: Aac, Eac, Iac, Oac, Aca, Eca, Ica, Oca, and NVC. We consider two lists: a list for WCS predictions and a list for participants' answers. In the WCS list, we assign 1 to an answer if it is predicted under WCS; else we assign 0. In the list with the participants' answers we use a threshold function and assign 1 to answers that were drawn by more than 16% of the participants; else we assign 0. Both lists can then be compared for their congruence as follows, where i is the ith element of both lists:

$$\text{COMP}(i) = \begin{cases} 1 & \text{if both have the same value for } i\text{th element,} \\ 0 & \text{otherwise.} \end{cases}$$

The matching percentage of this syllogism is then computed by $\sum_{i=1}^{9} \text{COMP}(i)/9$. Note that the percentage of the match does not only take into account when WCS correctly predicts a conclusion, but also whenever it correctly rejected a conclusion. The average percentage of accuracy is then the average of the matching percentage of all 64 syllogisms.

OA4 - Perfect Match (100%)

Consider $\mathcal{P}_{\mathsf{OA4}}$. The least model of wc $\mathcal{P}_{\mathsf{OA4}}$, $I = \langle I^\top, I^\bot \rangle$, is:

$$I^\top = \{b(o_1),\ b(o_2),\ b(o_3),\ ab_{ca}(o_1),\ a'(o_1),\ c(o_1),\ c(o_2),\ c(o_3)\},$$
$$I^\bot = \{ab_{bna}(o_1),\ ab_{naa}(o_1),\ ab_{bc}(o_1),\ ab_{bc}(o_2),\ ab_{bc}(o_3),\ a(o_1)\}.$$

This model entails only the conclusion '*some c are not a*' (Oca): There exists an object, viz. o_1, such that $\mathcal{P}_{\mathsf{OA4}} \models_{wcs} c(o_1) \wedge \neg a(o_1)$ and there exists another object, viz. o_2, such that $\mathcal{P}_{\mathsf{OA4}} \models_{wcs} c(o_2)$ and $\mathcal{P}_{\mathsf{OA4}} \not\models_{wcs} \neg a(o_2)$.

IE4 - Partial Match (89%)

Consider $\mathcal{P}_{\mathsf{IE4}}$. The least model of wc $\mathcal{P}_{\mathsf{IE4}}$, $I = \langle I^\top, I^\bot \rangle$, is

$$I^\top = \{a(o_i) \mid i \in \{1,3,4\}\} \cup \{b(o_i) \mid i \in \{1,2,3,5\}\} \cup \{c'(o_i) \mid i \in \{1,2,3,5\}\}$$
$$I^\bot = \{ab_{ba}(o_1), ab_{ab}(o_3), ab_{ncc}(o_5), ab_{bnc}(o_5)\} \cup \{ab_{bnc}(o_i) \mid i \in \{1,2,3,4,5\}\} \cup \{c(o_i) \mid i \in \{1,2,3,5\}\}.$$

This model entails only '*Some a are not c*' (Oac).

IA2 - Explain NVC: Perfect Match (100%)

Consider $\mathcal{P}_{\mathsf{IA2}}$. The least model of wc $\mathcal{P}_{\mathsf{IA2}}$, $I = \langle I^\top, I^\bot \rangle$, is

$$I^\top = \{a(o_1), a(o_3), a(o_4), b(o_1), b(o_2), b(o_3), b(o_5), c(o_5)\},$$
$$I^\bot = \{ab_{ba}(o_1), ab_{ab}(o_3)\} \cup \{ab_{cb}(o_i) \mid i \in \{1,2,3,4,5\}\}.$$

This model entails *'no valid conclusion'* (NVC). However, a significant percentage of participants answered Iac and Ica, despite IA2 being an invalid syllogism in classical FOL. According to the sixth principle, abduction, we believe that these participants might have searched for alternatives to NVC. We model this by applying skeptical abductive reasoning.

Each head of an existential import generates a single observation. We apply abduction sequentially to each of them. To prevent empty explanations we remove from the current program the fact that generated the observation. For each observation and each of its minimal explanations we compute the least model of the weak completion of the program extended with the explanation and collect all entailed syllogistic conclusions. Observations that cannot be explained are filtered out. The set Answers consists of all entailed conclusions for the observations left. For the final conclusions, we apply skeptical reasoning, i.e., the final answer to the current syllogism is given by $\mathsf{FinalAnswer} = \bigcap_{A \in \mathsf{Answers}} A$. In the case that FinalAnswer is empty, we entail the NVC conclusion.

Reconsider IA2, where the observations are $O_1 = \{b(o_1)\}$, $O_2 = \{a(o_3)\}$ and $O_3 = \{c(o_5)\}$. If we examine $O_i = \{\ o\ \}$ with $i \in \{1,2,3\}$, then we will try to find an explanation for O_i with respect to $\mathcal{P}_{\mathsf{IA2}} \setminus \{o \leftarrow \top\}$.[7] The set of abducibles is:

$$\begin{aligned} \mathcal{A}_{\mathcal{P}_{\mathsf{IA2}}} &= \{ab_{ba}(o_i) \leftarrow \top,\ ab_{ba}(o_i) \leftarrow \bot \mid i \in \{2,3,4,5\}\} \\ &\cup \{ab_{ab}(o_i) \leftarrow \top,\ ab_{ab}(o_i) \leftarrow \bot \mid i \in \{1,2,4,5\}\} \\ &\cup \{c(o_i) \leftarrow \top,\ \ c(o_i) \leftarrow \bot \mid i \in \{1,2,3,4\}\} \\ &\cup \{ab_{cb}(o_5) \leftarrow \top \mid i \in \{1,2,3,4,5\}\} \\ &\cup \{ab_{ba}(o_1) \leftarrow \top, ab_{ab}(o_3) \leftarrow \top\ \}. \end{aligned}$$

$\mathcal{E}_1 = \{c(o_1) \leftarrow \top\}$ and $\mathcal{E}_2 = \{c(o_3) \leftarrow \top, ab_{ba}(o_3) \leftarrow \bot\}$ are the minimal explanations for O_1 and O_2, respectively. Note that for O_3 there is no explanation.

Consider the observation $O_1 = \{b(o_1)\}$ and the program $\mathcal{P}^1_{\mathsf{IA2}} = (\mathcal{P}_{\mathsf{IA2}} \setminus \{b(o_1) \leftarrow \top\}) \cup \mathcal{E}_1$. The least model of wc $\mathcal{P}^1_{\mathsf{IA2}}$ is $\langle I^\top \cup \{c(o_1)\}, I^\bot \rangle$ where $\langle I^\top, I^\bot \rangle$ is the least model of wc $\mathcal{P}_{\mathsf{IA2}}$, as defined before. Thus, $c(o_1)$ is newly entailed to be true after applying abduction. This model entails what participants concluded, namely Iac and Ica. Iac is entailed as there exists an object, viz. o_1, such that $\mathcal{P}^1_{\mathsf{IA2}} \models_{wcs} a(o_1) \wedge c(o_1)$ and there exists another object, viz. o_4, such that $\mathcal{P}^1_{\mathsf{IA2}} \models_{wcs} a(o_4)$ and $\mathcal{P}^1_{\mathsf{IA2}} \not\models_{wcs} c(o_4)$, and there exists another object ,viz. o_5, such that $\mathcal{P}^1_{\mathsf{IA2}} \models_{wcs} c(o_5)$ and $\mathcal{P}^1_{\mathsf{IA2}} \not\models_{wcs} a(o_5)$. Analogously, *'some c are a'* (Ica) holds.

For the observation $O_2 = \{a(o_3)\}$ we consider the program $\mathcal{P}^2_{\mathsf{IA2}} = (\mathcal{P}_{\mathsf{IA2}} \setminus \{a(o_3) \leftarrow \top\}) \cup \mathcal{E}_2$. The least model of $\mathcal{P}^2_{\mathsf{IA2}}$ also entails the conclusions Iac and Ica.

$\mathsf{Answers}(\mathcal{P}_{\mathsf{IA2}}) = \{\{\mathsf{I}ac, \mathsf{I}ca\}, \{\mathsf{I}ac, \mathsf{I}ca\}\}$ is the collection of all conclusions. $\mathsf{FinalAnswer}(\mathcal{P}_{\mathsf{IA2}}) = \{\mathsf{I}ac, \mathsf{I}ca\}$ consists of the skeptically entailed conclusions, i.e. it is the intersection of all conclusions, which in this case are '*some a are c*' (Iac) and '*some c are a*' (Ica).

Overall Accuracy of 89%

The results of the three examples formalized under WCS are summarized and compared to FOL, PSYCOP, the Verbal, and the Mental Model Theory in Table 5. For some syllogisms the conclusions drawn by the participants and WCS are identical and for some syllogisms the conclusions drawn by the participants and WCS overlap. Combining the syllogistic premises representation and entailment rules for all 64 syllogistic premises and applying abduction when NVC was entailed (which happened in 43 cases), we accomplished an average of 89% accuracy in our predictions. In 18 cases we have a perfect match, in 30 cases the match is 89%, in 13

[7] We remove the fact from the program that generated the observation, because otherwise the explanation would be empty.

Table 5: The conclusions drawn by a significant percentage of participants are highlighted in gray and compared to the predictions of the theories FOL, PSYCOP, Verbal, and Mental Models as well as WCS for the syllogisms OA4, IE4, and IA2.

	Part.	FOL	PSYCOP	Verbal Models	Mental Models	WCS
OA4	Oca	Oca	Oca, Ica, Iac	Oca, NVC	Oca, Oac, NVC	Oca
IE4	Oac, NVC	Oac	Oac, Iac, Ica	Oac, NVC	Oac, NVC Eac, Eca, Oca	Oac
IA2	Iac, Ica	NVC	NVC	Ica NVC	Iac, Ica, NVC	Iac Ica

cases the match is 78%, and in the remaining three cases the match is 67%. We achieve the best performance compared to the other state-of-the-art cognitive theories with the current best performance of 84 % (Verbal Model Theory).

Conclusions

We developed seven principles for modeling a logical form for the representation of quantified statements in human reasoning, mostly motivated from findings in Cognitive Science. We show how these principles can be encoded within a computational logic approach, the Weak Completion Semantics. After that we discuss the predictions of three examples under WCS and compare them to the conclusions humans draw from in (Khemlani & Johnson-Laird, 2012). The result with respect to all 64 syllogistic premises under WCS shows that we achieve the best results with a prediction of 89%, compared to the results of other cognitive theories.

References

Dietz, E.-A. (2017). A computational logic approach to the belief bias in human syllogistic reasoning. In P. Brézillon, R. Turner, & C. Penco (Eds.), *10th int. and interdisciplinary conference on modeling and using context* LNAI (Vol. 10257, p. 691-707). Springer.

Dietz, E.-A., & Hölldobler, S. (2015). A new computational logic approach to reason with conditionals. In F. Calimeri, G. Ianni, & M. Truszczynski (Eds.), *Logic programming and nonmonotonic reasoning - 13th int. conference,* LNAI (Vol. 9345, pp. 265–278). Springer.

Dietz, E.-A., Hölldobler, S., & Höps, R. (2015). A computational logic approach to human spatial reasoning. In *IEEE symposium series on computational intelligence* (pp. 1627–1634). IEEE.

Dietz, E.-A., Hölldobler, S., & Pereira, L. M. (2015). On conditionals. In G. Gottlob, G. Sutcliffe, & A. Voronkov (Eds.), *Global conference on artificial intelligence,* Epic Series in Computing, 36 (pp. 79–92). EasyChair.

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

Dietz, E.-A., Hölldobler, S., & Ragni, M. (2013). A computational logic approach to the abstract and the social case of the selection task. In *11th int. symposium on logical formalizations of commonsense reasoning.*

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3). Academic Press.

Hölldober, S. (2009). *Logik und Logikprogrammierung* (Vol. 1: Grundlagen). Synchron Publishers.

Hölldobler, S. (2015). Weak completion semantics and its applications in human reasoning. In U. Furbach & C. Schon (Eds.), *CEUR WS proc. on bridging the gap between human and automated reasoning* (pp. 2–16).

Hölldobler, S., & Kencana Ramli, C. D. (2009a). Logic programs under three-valued Łukasiewicz semantics. In P. M. Hill & D. S. Warren (Eds.), *25th int. conference on logic programming,* LNCS 5649 (pp. 464–478). Springer.

Hölldobler, S., & Kencana Ramli, C. D. (2009b). Logics and networks for human reasoning. In *int. conference on artificial neural networks,* LNCS, 5769 (pp. 85–94). Springer.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 427-457.

Lloyd, J. W. (1984). *Foundations of logic programming*. Springer.

Łukasiewicz, J. (1920). O logice trójwartościowej. *Ruch Filozoficzny*, *5*, 169–171.

Peirce, C., Hartshorne, C., & Weiss, P. (1974). *Collected papers of charles sanders peirce*. Belknap Press of Harvard University Press.

Pereira, L. M., Dietz, E.-A., & Hölldobler, S. (2014a). An abductive reasoning approach to the belief-bias effect. In C. Baral, G. D. Giacomo, & T. Eiter (Eds.), *Principles of knowledge representation and reasoning: Proc. of the 14th int. conference* (p. 653-656). AAAI Press.

Pereira, L. M., Dietz, E.-A., & Hölldobler, S. (2014b). Contextual abductive reasoning with side-effects. In I. Niemelä (Ed.), *Theory and practice of logic programming,* 14 (p. 633-648). Cambridge University Press.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, *102*(3), 533.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. MIT Press.

Smith, R. (1989). *Aristotle's prior analytics*. Hackett.

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