Principles and Clusters in Human Syllogistic Reasoning

Emmanuelle-Anna Dietz Saldanha, Steffen Hölldobler, and Richard Mörbitz[⋆]

International Center for Computational Logic, TU Dresden, Germany,
{dietz,sh}@iccl.tu-dresden.de, richard.moerbitz@tu-dresden.de

Abstract. It seems widely accepted that human reasoning cannot be modeled by means of Classical Logic. Psychological experiments have repeatedly shown that participants' answers systematically deviate from the classical logically correct answers. Recently a new approach on modeling human syllogistic reasoning has been developed which seems to perform the best compared to other state-of-the-art cognitive theories. We take this approach as starting point, yet instead of trying to model *the* human reasoner, we aim at identifying clusters of reasoners, which can be characterized by principles or by heuristic strategies.

1 Introduction

In recent years, a new cognitive theory based on the Weak Completion Semantics (WCS) has been developed. It has its roots in the ideas first expressed by Stenning and van Lambalgen [19], but is mathematically sound [8], and has been successfully applied to various human reasoning tasks. An overview can be found in [7]. Hence, it was natural to ask whether WCS is competitive in syllogistic reasoning and how it performs wrt the cognitive theories evaluated in [12]. Consider the following quantified statements:

$$\textit{All a are b.Some c are not b.} \tag{AO3}$$

Classical logically *Some c are not a* follows from these premises. However, according to [12], the majority of participants in experimental studies, concluded that *no valid conclusion* and *Some c are not a* follows. Yet, these two responses exclude each other, i.e. it is unlikely that the participants who answered *no valid conclusion* are the same ones who answered *Some c are not a*, and vice versa.

The four quantifiers and their formalization in FOL are given in Table 1. The entities can appear in four different orders called *figures* 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 AO3.

Recently, a computational logic approach to human syllogistic reasoning has been developed under the Weak Completion Semantics, which identifies seven principles for modeling the logical form of the representation of quantified statements in human reasoning [1]. The results of this approach achieved a match

[⋆] The authors are mentioned in alphabetical order.

Mood	First-order logic	Short
affirmative universal	$\forall X(a(X) \rightarrow b(X))$	Aab
affirmative existential	$\exists X(a(X) \wedge b(X))$	Iab
negative universal	$\forall X(a(X) \rightarrow \neg b(X))$	Eab
negative existential	$\exists X(a(X) \wedge \neg b(X))$	Oab

Table 1. The moods and their formalization.

	1st Premise	2nd Premise
Fig. 1	a-b	b-c
Fig. 2	b-a	c-b
Fig. 3	a-b	c-b
Fig. 4	b-a	b-c

Table 2. The four figures.

of 89% with respect to the conclusions participants gave, based on the data reported in [12]. This result stands out because the best of the twelve other state-of-the-art cognitive theories, only achieved a match of 84%.

While reasoning with conditionals humans seems to take certain assumptions for granted which however are not stated explicitly in the task description. As psychological experiments show, these assumptions seem not to be arbitrary but instead are systematic in the sense that they are repeatedly made by participants. Furthermore, some assumptions reappear in various experiments, whereas other assumptions are only made in very few experiments or only by some participants. In order to identify and structure these assumptions, we view them as principles that are either applied or ignored by the participants who have to solve the task. As starting point, we take the syllogistic reasoning approach presented in [1]. However, a major drawback of this approach is that only the matching with respect to the aggregated data is considered, i.e. the approach models *the* human reasoner. However, the above example and other examples such as cases of the Wason Selection Task reported in [15], serve as indication that *the* human reasoner does not exist, but instead we might better search for clusters of human reasoners. These clusters might be expressed by principles, i.e. some clusters might apply some principles that are not applied by other clusters.

The paper is structured as follows: First, we present the principles for the representation of quantified statements, motivated by findings from Cognitive Science and Linguistics. Next, the Weak Completion Semantics is introduced and the encoding of quantified statements within this approach in Section 3 and 4. Then the clusters and heuristics are discussed and finally an overall evaluation of the Weak Completion Semantics is presented.

2 Principles about Quantified Statements

Eight principles for developing a logical form of quantified statements are presented. They originate from [1,2] except of the principles in Section 2.5 and 2.8.

2.1 Quantified Statements as Implication (conditionals)

Independent of the quantifiers mood, we decide to formalize any relation between two objects of a quantified statement by means of the implication such that the first object is the antecedent and the second object the conclusion in the implication. For instance, the statement *All a are b* is expressed as $\forall X(a(X) \rightarrow b(X))$.

2.2 Licenses for Inferences (licenses)

[19] proposed to formalize conditionals in human reasoning not by inferences straight away, but rather by *licenses for inferences.* Given the quantified statement *All a are b*, a license for this inference can then be expressed by *All a that are not abnormal, are b*. Given the previous formalization of this statement as $\forall X(a(X) \rightarrow b(X))$, we extend this implication by conjoining $a(X)$ together with an abnormality predicate as follows: $\forall X(a(X) \wedge \neg ab_{pq}(X) \rightarrow b(X))$. Further, the closed-world assumption with respect to the abnormality predicate is expressed by *nothing is abnormal wrt* X, i.e. $\neg ab_{pq}(X)$.

2.3 Existential Import and Gricean Implicature (import)

Humans understand quantifiers differently due to a pragmatic understanding of the 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* [6]. This corresponds to what sometimes in literature is also called *existential import* and assumed by several theories like the theory of mental models [11] or mental logic [18]. Likewise, [19] have shown that humans require existential import for a conditional to be true.

Furthermore, as mentioned by [12], the quantifier *some a are b* often implies that *some a are not b*, which again is implied by 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 *some a are b* instead, we assume that *not all a are b*, which in turn implies *some a are not b.*

2.4 Unknown Generalization (unknownGen)

Humans seem to distinguish between *some y are z* and *some z are y*, as the results reported by [12] show. Nevertheless, if we would represent *some y are z* by $\exists X(y(X) \wedge z(X))$ then this is semantically equivalent to $\exists X(z(X) \wedge y(X))$ because conjunction is commutative in FOL. Likewise, humans seem to distinguish between *some y are z* and *all y are z*, as we have already discussed in Section 2.3. Accordingly, if we only 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 distinguish between *some y are z* and *all y are z*, 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 but there must be another object o_2, which belongs to y and for which it is unknown whether it belongs to z. To express this idea, we can make use of the the principle (licenses) presented in Section 2.2 as follows: We replace $\neg ab_{pq}(X)$ by $\neg ab_{pq}(o_1)$, i.e. the closed-world assumption about abnormal is only applied wrt o_1.

2.5 Deliberate Generalization (deliberateGen)

If all of the principles introduced so far are applied to an existential premise, the only object about which an inference can be made is the one resulting from the existential import principle. This is because the abnormality introduced by the licenses for inferences principle has to be false for inference, but due to the unknown generalization principle it is unknown for other objects.

There is, however, evidence that some humans still draw conclusions in such circumstances [12]. We believe that they do not take into account abnormalities regarding objects that are not related to the premise.

2.6 Converse Implication (converse)

Although there seems to be some evidence that humans distinguish between *some y are z* and *some z are y* (see the results reported in [12]) 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.

2.7 Search Alternative Conclusions to NVC (searchAlt)

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 about the facts coming 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.

2.8 Contraposition (contraposition)

In FOL, a conditional statement of the form $\forall(X)(a(X) \leftarrow b(X))$ is logically equivalent to its *contrapositive* $\forall(X)(\neg b(X) \leftarrow \neg a(X))$. This contraposition also holds for the syllogistic moods A and E. There is evidence in [12] that some of the participants make use of this equivalence when solving syllogistic reasoning tasks. We believe that when they encounter a premise with the mood A (e.g. *All a are b*), then they might reason with the contrapositive conditional as well.

3 Weak Completion Semantics

The general notation, which we will use in the paper, is based on [13].

3.1 Contextual Logic Programs

Contextual logic programs are (data) logic programs extended by the truth-functional operator ctxt, called *context* [5]. *(Propositional) contextual logic program clauses* are expressions of the forms $A \leftarrow L_1 \wedge \ldots \wedge L_m \wedge \mathsf{ctxt}(L_{m+1}) \wedge$

$\ldots \wedge \mathsf{ctxt}(L_{m+p})$ (called *rules*), $A \leftarrow \top$ (called *facts*), $A \leftarrow \bot$ (called *negative assumptions*) and $A \leftarrow \mathsf{U}$ (called *unknown assumptions*). A is an atom and the L_i with $1 \leq i \leq m+p$ are literals. A is called *head* and $L_1 \wedge \ldots \wedge L_m \wedge \mathsf{ctxt}(L_{m+1}) \wedge \ldots \wedge \mathsf{ctxt}(L_{m+p})$ as well as $\top, \bot$ and U, standing for *true*, *false* and *unknown* respectively, are called *body* of the corresponding clauses. A *contextual (logic) program* is a set of contextual logic program clauses. $\mathsf{g}\mathcal{P}$ denotes the set of all ground instances of clauses occurring in $\mathcal{P}$. $\mathsf{atoms}(\mathcal{P})$ denotes the set of all atoms occurring in $\mathsf{g}\mathcal{P}$. A is *defined in* $\mathcal{P}$ iff $\mathcal{P}$ contains a rule or a fact with head A. A is *undefined in* $\mathcal{P}$ iff A is not defined in $\mathcal{P}$. The set of all atoms that are undefined in $\mathcal{P}$ is denoted by $\mathsf{undef}(\mathcal{P})$. The *definition of* A *in* $\mathcal{P}$ is defined as $\mathit{def}(A, \mathcal{P}) = \{A \leftarrow \mathit{Body} \mid A \leftarrow \mathit{Body}$ is a rule or a fact occurring in $\mathcal{P}\}$. $\neg A$ is *negatively assumed in* $\mathcal{P}$ iff $\mathcal{P}$ contains an negative assumption with head A, no unknown assumption with head A and $\mathit{def}(A, \mathcal{P}) = \emptyset$. We omit the word *contextual* when we refer to programs, if not stated otherwise.

3.2 Integrity Constraints

A set of *integrity constraints* IC consists of clauses of the form $\mathsf{U} \leftarrow \mathit{Body}$, where *Body* is a conjunction of literals and U denotes the unknown. Hence, an interpretation maps an integrity constraint to $\top$ iff *Body* is either mapped to U or $\bot$. This understanding is similar to the definition of the integrity constraints for the Well-founded Semantics in [14]. Given an interpretation I and a set of integrity constraints IC, I *satisfies* IC iff all clauses in IC are true under I.

3.3 Three-Valued Łukasiewicz Logic Extended by ctxt Connective

We consider the three-valued Łukasiewicz logic together with the ctxt connective, for which the corresponding truth values are $\top$, $\bot$ and U, meaning *true*, *false* and *unknown*, respectively. A *three-valued interpretation* I is a mapping from $\mathsf{atoms}(\mathcal{P})$ to the set of truth values $\{\top, \bot, \mathsf{U}\}$, represented as a pair $I = \langle I^\top, I^\bot \rangle$ of two disjoint sets of atoms: $I^\top = \{A \mid A$ is mapped to $\top$ under $I\}$ and $I^\bot = \{A \mid A$ is mapped to $\bot$ under $I\}$. Atoms which do not occur in $I^\top \cup I^\bot$ are mapped to U. The truth value of a given formula under I is determined according to the truth tables in Table 3. $I(F) = \top$ means that a formula F is mapped to true under I. A *three-valued model* $\mathcal{M}$ of $\mathcal{P}$ is a three-valued interpretation such that $\mathcal{M}(A \leftarrow \mathit{Body}) = \top$ for each $A \leftarrow \mathit{Body} \in \mathcal{P}$. 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 is the *least model* of $\mathcal{P}$ iff for any other model J of $\mathcal{P}$ it holds that $I \subseteq J$.

3.4 Forward Reasoning: Least Models under the Weak Completion

For a given $\mathcal{P}$, consider the following transformation: 1.For each ground atom A which is defined in $\mathcal{P}$, replace all clauses of the form $A \leftarrow \mathit{Body}_1, \ldots, A \leftarrow \mathit{Body}_m$ occurring in $\mathsf{g}\mathcal{P}$ by $A \leftarrow \mathit{Body}_1 \vee \ldots \vee \mathit{Body}_m$. 2. Replace all occurrences of $\leftarrow$ by $\leftrightarrow$. The obtained ground program is called *weak completion* of $\mathcal{P}$ or $\mathit{wc}\mathcal{P}$.

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	$\bot$

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

$\leftrightarrow$	$\top$	U	$\bot$
$\top$	$\top$	U	$\bot$
U	U	$\top$	U
$\bot$	$\bot$	U	$\top$

L	$\mathsf{ctxt}(L)$
$\top$	$\top$
$\bot$	$\bot$
U	$\bot$

Table 3. The truth tables for the connectives under the three-valued Łukasiewicz logic and for $\mathsf{ctxt}(L)$. L is a literal, $\top$, $\bot$, and U denote *true*, *false*, and *unknown*, respectively.

Consider the following semantic operator, which is due to Stenning and van Lambalgen [19]: Let $I = \langle I^\top, I^\bot \rangle$ be an interpretation. $\Phi_\mathcal{P}(I) = \langle J^\top, J^\bot \rangle$, where

$$\begin{aligned} J^\top &= \{A \mid A \leftarrow \mathit{Body} \in \mathit{def}(A,\mathcal{P}) \text{ and } \mathit{Body} \text{ is } \mathit{true} \text{ under } \langle I^\top, I^\bot \rangle\} \\ J^\bot &= \{A \mid \mathit{def}(A,\mathcal{P}) \neq \emptyset \text{ and} \\ &\qquad \mathit{Body} \text{ is } \mathit{false} \text{ under } \langle I^\top, I^\bot \rangle \text{ for all } A \leftarrow \mathit{Body} \in \mathit{def}(A,\mathcal{P})\} \end{aligned}$$

The least fixed point of $\Phi_\mathcal{P}$ is denoted by $\mathsf{lfp}\,\Phi_\mathcal{P}$, if it exists. [9] showed that non-contextual programs as well as their weak completions always have a least model under Łukasiewicz logic, which can be obtained as the least fixed point of Φ. However, for programs with the ctxt operator this property only holds if the programs do not contain cycles [5]. We define $\mathcal{P} \models_{wcs} F$ iff $\mathcal{P}$ is acyclic and $\mathsf{lfp}\,\Phi_\mathcal{P} \models F$. In the remainder of this paper, we only consider acyclic programs and $\mathcal{M}_\mathcal{P}$ denotes the least fixed point of $\Phi_\mathcal{P}$.

3.5 Backward Reasoning: Explanations by Means of Abduction

An *abductive framework* $\langle \mathcal{P}, \mathcal{A}, \mathsf{IC}, \models_{wcs} \rangle$ consists of a program $\mathcal{P}$, a set $\mathcal{A}$ of abducibles, a set IC of integrity constraints, and the entailment relation $\models_{wcs}$. The set of abducibles $\mathcal{A} = \{A \leftarrow \top \mid \mathit{def}(A,\mathcal{P}) = \emptyset\} \cup \{A \leftarrow \bot \mid A \in \mathsf{undef}(\mathcal{P})\}$. Let $\langle \mathcal{P}, \mathcal{A}, \mathsf{IC}, \models_{wcs} \rangle$ be an abductive framework and *observation* $\mathcal{O}$ a set of literals. $\mathcal{O}$ is *explainable* in $\langle \mathcal{P}, \mathcal{A}, \mathsf{IC}, \models_{wcs} \rangle$ if and only if there exists an $\mathcal{E} \subseteq \mathcal{A}$, such that $\mathcal{P} \cup \mathcal{E} \models L$ for all $L \in \mathcal{O}$ and $\mathcal{P} \cup \mathcal{E}$ satisfies $\mathcal{IC}$. $\mathcal{E}$ is then called *explanation* for $\mathcal{O}$ given $\mathcal{P}$ and IC. We restrict $\mathcal{E}$ to be *minimal*, i.e. there does not exist any other explanation $\mathcal{E}' \subseteq \mathcal{A}$ for $\mathcal{O}$ such that $\mathcal{E}' \subseteq \mathcal{E}$.

Among the minimal explanations, it is possible that some of them entail a certain formula F while others do not. There exist two strategies to determine whether F is a valid conclusion in such cases. F follows *credulously*, if it is entailed by at least one explanation. F follows *skeptically*, if it is entailed by all explanations. Due to previous results on modeling human reasoning [3,4,1], skeptical abduction is applied. The set of observations wrt $\mathcal{P}$, $\mathcal{O}_\mathcal{P}$, as follows:

$$\mathcal{O}_\mathcal{P} = \{\{A\} \mid A \leftarrow \top \in def(A,\mathcal{P}) \wedge (A \leftarrow B_1 \wedge \cdots \wedge B_n) \in def(A,\mathcal{P})\},$$

where $n > 0$ and B_i is a literal for all $1 \leq i \leq n$. These are the atoms that occur in the head of a both rule and a fact. In the following, the idea is find an explanation for each observation $\mathcal{O} \in \mathcal{O}_\mathcal{P}$ where the observation is further restricted by considering only facts that result from certain principles.

3.6 Encoding Aspects about Quantified Statements

Negation by Transformation (transformation) The logic programs we consider do not allow heads of clauses to be negative literals. A negative conclusion $\neg p(X)$ is represented by introducing an auxiliary formula $p'(X)$ together with the 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 (licenses) introduced in Section 2.2, this additional clause is extended by the following two clauses: $p(X) \leftarrow \neg p'(X) \wedge \neg ab_{npp}(X)$. $ab_{npp}(X) \leftarrow \bot$. Additionally, the integrity constraint $\mathsf{U} \leftarrow p(X) \wedge p'(X)$ states that an object cannot belong to both, p and p'.

No Derivation through Double Negation (doubleNeg) A positive conclusion can be derived from double negation within two conditionals. Consider the following two conditionals with each one having a negative premise: If not a, then b. If not b then c. Additionally, assume that a is true. Let us encode the two conditionals and the fact that a is true as $\mathcal{P} = \{b \leftarrow \neg a, c \leftarrow \neg b, a \leftarrow \top\}$. $\mathsf{wc}\,\mathcal{P}$ is $\{b \leftrightarrow \neg a, c \leftrightarrow \neg b, a \leftrightarrow \top\}$ where $\mathcal{M}_{\mathcal{P}} = \langle\{a, c\}, \{b\}\rangle \models a \wedge c$. It appears to be the case that humans do not reason in such a way, considering the results of the participants' responses in [12]. Accordingly, we block them through abnormalities.

4 Quantified Statements as Logic Programs

Based on the principles and encoding aspects in Section 2 and Section 3.6, we encode the quantified statements into logic programs. The programs are specified using the predicates y and z and depending on the figures shown in Table 2, where yz can be replaced by ab, ba, cb or bc. Here, all principles regarding a premise are described. However, we will later assume different clusters of reasoners, some of which do not apply certain principles (see Section 5). For such clusters, the clauses associated with the principles not applied are removed from the program.

4.1 All y are z (Ayz)

All y are z is represented by $\mathcal{P}_{Ayz}$, which consists of the following clauses:

$$
\begin{array}{rll}
z(X) & \leftarrow y(X) \wedge \neg ab_{yz}(X). & \text{(conditionals\&licenses)} \\
ab_{yz}(X) & \leftarrow \bot. & \text{(licenses)} \\
y(o) & \leftarrow \top. & \text{(import)} \\
ab_{yz}(X) & \leftarrow \mathsf{ctxt}(z'(X)). & \text{(contraposition \& licenses \& deliberateGen)} \\
y'(X) & \leftarrow \neg z(X) \wedge \neg ab_{zy}(X). & \text{(contraposition \& conditionals \& licenses)} \\
ab_{zy}(X) & \leftarrow \bot. & \text{(contraposition \& licenses)} \\
y(X) & \leftarrow \neg y'(X) \wedge \neg ab_{nyy}(X). & \text{(contraposition \& transformation\&licenses)}
\end{array}
$$

The first two clauses are obtained by applying the principles of representing quantified statements as implication and licenses for inferences. The third clause follows by the principle of existential import and Gricean implicature. The last four clauses result from applying the contraposition principle. The deliberate generalization principle must also be used, because otherwise inference of $\neg z(X)$ would not be possible. It defeats the original assumption $ab_{yz}(X) \leftarrow \bot$ in the sense that the weak completion of

$$ab_{yz}(X) \leftarrow \bot, ab_{yz}(X) \leftarrow \mathsf{ctxt}(z'(X))$$

is $ab_{yz}(X) \leftrightarrow \bot \vee \mathsf{ctxt}(z'(X))$, which is equivalent to $ab_{yz}(X) \leftrightarrow \mathsf{ctxt}(z'(X))$. As the contrapositive conditional would have a negative atom in the head, the negation by transformation encoding is used. Note that there is no import of an object for which $\neg z(X)$ holds, because this does not follow from the premises. Consequently, the abnormality introduced by the principles licenses for inferences and negation by transformation does not have to be assumed as false for any object. $\mathcal{M}_{\mathcal{P}_{Ayz}}$ is $\langle \{y(o), z(o)\}, \{ab_{yz}(o)\} \rangle$. If contraposition is applied, by negation by transformation we have the following integrity constraint: $\mathsf{U} \leftarrow y(X) \wedge y'(X)$.

4.2 No y is z (Eyz)

No y is z is represented by $\mathcal{P}_{Eyz}$, which consists of the following clauses:

$z'(X) \leftarrow y(X) \wedge \neg ab_{ynz}(X)$.	(transformation & licenses)
$ab_{ynz}(X) \leftarrow \bot$.	(licenses)
$z(X) \leftarrow \neg z'(X) \wedge \neg ab_{nzz}(X)$.	(transformation & licenses)
$y(o_1) \leftarrow \top$.	(import)
$ab_{nzz}(o_1) \leftarrow \bot$.	(licenses & doubleNeg)
$y'(X) \leftarrow z(X) \wedge \neg ab_{zny}(X)$.	(converse & transformation & licenses)
$ab_{zny}(X) \leftarrow \bot$.	(converse&licenses)
$y(X) \leftarrow \neg y'(X) \wedge \neg ab_{nyy}(X)$.	(converse & transformation & licenses)
$z(o_2) \leftarrow \top$.	(converse&import)
$ab_{nyy}(o_2) \leftarrow \bot$.	(converse & licenses & doubleNeg)

In addition, we have the following two integrity constraints:

$\mathsf{U} \leftarrow z(X) \wedge z'(X)$.	(transformation)
$\mathsf{U} \leftarrow y(X) \wedge y'(X)$.	(converse & transformation)

The first two clauses in $\mathcal{P}_{Eyz}$ are obtained by applying the principles of representing quantified statements as conditionals and using licenses for inferences, where $z'(X)$ is an auxiliary formula used to denote the negation of $z(X)$. $z'(X)$ is related to $z(X)$ by the third clause applying negation by transformation. In addition, this principle enforces the integrity constraint. The fourth clause of $\mathcal{P}_{Eyz}$ follows by the principle of Gricean implicature and the fifth because of licenses for inferences and no derivation through double negation. The last five clauses are obtained by the same reasons as the first five clauses together with the principle of converse implication. Note that the last clause in $\mathcal{P}_{Eyz}$ cannot be generalized to all X, because otherwise we allow conclusions by double negatives. Therefore we apply the encoding doubleNeg. $\mathcal{M}_{\mathcal{P}_{Eyz}}$ is

$$\langle \{y(o_1), z'(o_1), z(o_2), y'(o_2)\}, \\ \{ab_{ynz}(o_1), ab_{nzz}(o_1), z(o_1), ab_{zny}(o_2), ab_{nyy}(o_2), y(o_2)\} \rangle.$$

4.3 Some y are z (Iyz)

Some y are z is represented by $\mathcal{P}_{Iyz}$, which consists of the following clauses:

$$
\begin{array}{rll}
z(X) & \leftarrow y(X) \wedge \neg ab_{yz}(X). & \text{(conditionals \& licenses)} \\
ab_{yz}(o_1) & \leftarrow \bot. & \text{(unknownGen \& licenses)} \\
y(o_1) & \leftarrow \top. & \text{(import)} \\
y(o_2) & \leftarrow \top. & \text{(unknownGen)} \\
ab_{yz}(X) & \leftarrow \mathsf{ctxt}(z'(X)). & \text{(licenses \& deliberateGen)} \\
ab_{yz}(o_2) & \leftarrow \mathsf{U}. & \text{(licenses \& deliberateGen)} \\
y(X) & \leftarrow z(X) \wedge \neg ab_{zy}(X). & \text{(converse \& conditionals\& licenses)} \\
ab_{zy}(o_3) & \leftarrow \bot. & \text{(converse\& licenses \& unknownGen)} \\
z(o_3) & \leftarrow \top. & \text{(converse \& import)} \\
z(o_4) & \leftarrow \top. & \text{(converse \& unknownGen)} \\
ab_{zy}(X) & \leftarrow \mathsf{ctxt}(y'(X)). & \text{(converse \& licenses \& deliberateGen)} \\
ab_{zy}(o_4) & \leftarrow \mathsf{U}. & \text{(converse \& licenses \& deliberateGen)}
\end{array}
$$

The first two clauses are again obtained by the principles of representing quantified statements as conditionals and using licenses for inferences. The abnormality predicate is restricted to the object o_1, which is assumed to exist by the principle of Gricean implicature, represented by the third clause. The fourth clause is obtained by the principle of unknown generalization. The fifth and sixth clause are obtained by the principle of unknown generalization. The last six clauses are obtained by the same reasons as the first six clauses together with the principle of converse implication. $\mathcal{M}_{\mathcal{P}_{Iyz}}$ is $\langle\{y(o_1), y(o_2), z(o_1)\}, \{ab_{yz}(o_1)\}\rangle$. Note $ab_{yz}(o_2)$ is an unknown assumption in $\mathcal{P}_{Iyz}$. Accordingly, $z(o_2)$ stays unknown in $\mathcal{M}_{\mathcal{P}_{Iyz}}$.

4.4 Some y are not z (Oyz)

Some y are not z is represented by $\mathcal{P}_{Oyz}$ which consists of the following clauses:

$$
\begin{array}{rll}
z'(X) & \leftarrow y(X) \wedge \neg ab_{ynz}(X). & \text{(conditionals \& transformation \& licenses)} \\
ab_{ynz}(o_1) & \leftarrow \bot. & \text{(unknownGen \& licenses)} \\
z(X) & \leftarrow \neg z'(X) \wedge \neg ab_{nzz}(X). & \text{(transformation \& licenses)} \\
y(o_1) & \leftarrow \top. & \text{(import)} \\
y(o_2) & \leftarrow \top. & \text{(unknownGen)} \\
ab_{nzz}(o_1) & \leftarrow \bot. & \text{(doubleNeg \& licenses)} \\
ab_{nzz}(o_2) & \leftarrow \bot. & \text{(doubleNeg \& licenses)}
\end{array}
$$

In addition, we have the following integrity constraint:

$$U \leftarrow z(X) \wedge z'(X). \qquad \text{(transformation)}$$

The first four clauses as well as the integrity constraint are derived as in the program $\mathcal{P}_{Eyz}$ except that object o_1 is used instead of o and ab_{ynz} is restricted to o_1 like in $\mathcal{P}_{Iyz}$. The fifth clause of $\mathcal{P}_{Oyz}$ is obtained by the principle of unknown generalization. The last two clauses are again not generalized to all objects for the same reason as previously discussed in Section 4.2 for the representation of E: The generalization of ab_{nzz} to all objects can lead to conclusions through double negation, in case there is a second premise. $\mathcal{M}_{\mathcal{P}_{Oyz}}$ is $\langle\{y(o_1), y(o_2), z'(o_1)\}, \{ab_{ynz}(o_1), ab_{nzz}(o_1), ab_{nzz}(o_2), z(o_1)\}\rangle$.

4.5 Entailment of Syllogisms

We define when $\mathcal{M}_\mathcal{P}$ entails a conclusion, where yz is to be replaced by ac or ca.

All (A) $\mathcal{P} \models Ayz$ iff there exists an object o such that $\mathcal{P} \models_{wcs} y(o)$ and for all objects o we find that if $\mathcal{P} \models_{wcs} y(o)$ then $\mathcal{P} \models_{wcs} z(o)$.

No (E) $\mathcal{P} \models Eyz$ iff there exists an object o_1 such that $\mathcal{P} \models_{wcs} y(o_1)$ and for all objects o_1 we find that if $\mathcal{P} \models_{wcs} y(o_1)$ then $\mathcal{P} \models_{wcs} \neg z(o_1)$ and if there exists an object o_2 such that $\mathcal{P} \models_{wcs} z(o_2)$ and for all objects o_2 we find that if $\mathcal{P} \models_{wcs} z(o_2)$ then $\mathcal{P} \models_{wcs} \neg y(o_2)$.

Some (I) $\mathcal{P} \models Iyz$ 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 $\mathcal{P} \models_{wcs} z(o_3) \wedge y(o_3)$ and there exists an object o_4 such that $\mathcal{P} \models_{wcs} z(o_4)$ and $\mathcal{P} \not\models_{wcs} y(o_4)$.

Some Are Not (O) $\mathcal{P} \models Oyz$ 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 $\mathcal{P} \not\models_{wcs} \neg z(o_2)$.

NVC When no previous conclusion can be derived, no valid conclusion holds.

4.6 Accuracy of Predictions

We have nine different answer possibilities for each of the 64 syllogisms:

Aac, Eac, Iac, Oac, Aca, Eca, Ica, Oca and NVC.

For every syllogism, we define a list of length 9 for the predictions of the Weak Completion Semantics, where the first element represents Aac, the second element represents Eac, and so forth. When Aac is predicted under the Weak Completion Semantics for a given syllogism, then the value of the first element of this list is a 1, otherwise it is a 0, and the same holds for the other eight elements in the list. Analogously, for every syllogism we define a list of the participants' conclusions of length 9 containing either 1 or 0 for all nine answer possibilities, depending on whether the majority concluded Aac, Eac, and so forth. For each syllogism we compare each element of both lists as follows, where i is the ith element of both lists:

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

The matching percentage of this syllogism is then computed by $\sum_{i=1}^{9} \textsc{comp}(i)/9$. Note that the percentage of the match does not only take in account when the Weak Completion Semantics correctly predicts a conclusion, but also whenever it correctly rejected a conclusion.

5 Clusters and Heuristics

We consider clusters of human reasoners in terms of principles. Each cluster is a group of humans that applies the same principles. When identifying such clusters, e.g. by among the participants of [12], the principles used by a single cluster should lead to a significant answer for the syllogism in question. As the answers of all participants have been accumulated in the meta-analysis, the combined answers of all clusters should exactly correspond to the significant answers for that syllogism.

5.1 Basic Principles

Basic principles are assumed to be applied by all reasoners, regardless of any cluster. These are conditionals, licenses, import, and unknownGen. Note that they are not necessarily applicable to every syllogism: unknownGen may only be used for premises with an existential mood.

5.2 Advanced Principles and Clusters

Advanced principles are assumed to be used by not all humans, making them the starting point for clusters. Advanced principles considered in this paper are converse, deliberateGen, contraposition, and searchAlt, but there may exist more. When two individuals differ in the sense that one applies such a principle and the other one does not, we assume that they belong to different clusters.

As an example, consider the syllogism AO3 introduced in Section 1. According to the encoding described in Section 4, it is represented as the following logic program $\mathcal{P}_{\mathsf{AO3},\text{basic}}$ if only the basic principles are applied:

$$\begin{array}{lll} b(X) \leftarrow a(X) \wedge \neg ab_{ab}(X). & b'(X) \leftarrow c(X) \wedge \neg ab_{cnb}(X). & c(o_3) \leftarrow \top. \\ ab_{ab}(X) \leftarrow \bot. & c(o_2) \leftarrow \top. & ab_{nbb}(o_2) \leftarrow \bot. \\ a(o_1) \leftarrow \top. & b(X) \leftarrow \neg b'(X) \wedge \neg ab_{nbb}(X). & ab_{cnb}(o_2) \leftarrow \bot. \\ & & ab_{nbb}(o_3) \leftarrow \bot. \end{array}$$

$\mathcal{M}$ of $\mathcal{P}_{\mathsf{AO3},\text{basic}}$ is

$$\begin{array}{l} \langle\ \{a(o_1), b(o_1), \boxed{c(o_2)}, \boxed{c(o_3)}, b'(o_2)\}, \\ \quad \{ab_{ab}(o_1), ab_{ab}(o_2), ab_{ab}(o_3), ab_{cnb}(o_2), ab_{nbb}(o_2), ab_{nbb}(o_3)\}\rangle. \end{array}$$

NVC follows from this model. If additionally contraposition is used, then we consider the following program instead:

$$\begin{array}{rl} \mathcal{P}_{\mathsf{AO3},\mathsf{contraposition}} = & \mathcal{P}_{\mathsf{AO3},\text{basic}} \cup \{a'(X) \leftarrow \neg b(X) \wedge \neg ab_{ba}(X),\ ab_{ba}(X) \leftarrow \bot, \\ & a(X) \leftarrow \neg a'(X) \wedge \neg ab_{naa}(X),\ ab_{ab}(X) \leftarrow \mathsf{ctxt}(b'(X))\}. \end{array}$$

$\mathcal{M}$ of $\mathcal{P}_{\mathsf{AO3},\mathsf{contraposition}}$ is as follows:

$$\begin{array}{l} \langle\ \{a(o_1), ab_{ab}(o_2), b(o_1), \boxed{c(o_2)}, \boxed{c(o_3)}, a'(o_2), b'(o_2)\}, \\ \quad \{\boxed{a(o_2)}, ab_{ab}(o_1), ab_{ab}(o_3), ab_{cnb}(o_2), ab_{nba}(o_1), ab_{nba}(o_2), \\ \quad ab_{nba}(o_3), ab_{nbb}(o_2), ab_{nbb}(o_3), b(o_2), a'(o_1)\}\rangle. \end{array}$$

It entails the conclusion Oca. Let us assume there are two clusters of people whose reasoning process differs in the application of the contraposition principle. We unite the conclusions predicted for the clusters just as the answers of the

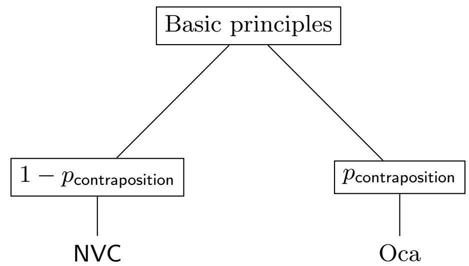

Fig. 1. MPT for the syllogism AO3.

participants of psychological studies are accumulated, obtaining $\{\mathsf{Oca}, \mathsf{NVC}\}$. These are exactly the significant answers reported in [12].

In order to represent what principles lead to what conclusions, Multinomial Processing Trees (MPTs) [17] are used. They have been suggested for modeling cognitive theories, because they represent cognitive processes as probabilistic procedures, thus being able to predict multiple answers and even their quantitative distribution [16]. We set the latent states (inner nodes) of the MPTs to the decisions whether to use certain principles and put the corresponding conclusions in the leaves. An MPT for the AO3 syllogism based on the clustering described above is presented in Figure 1. The parameter $p_{\mathsf{contraposition}}$ models the probability that an individual applies the contraposition principle and therefore belongs to the corresponding cluster. It can be trained from experimental data with algorithms like Expectation-Maximization [10]. Note that the MPT cannot predict all possible conclusions for a syllogism. This issue is addressed below.

5.3 Heuristic Strategies

Some theories suggest that some humans do not use logic at all to solve a syllogism, but rely on heuristics such as the atmosphere bias [21] or the matching bias [20]. Given the participants' answers presented in [12], it seems that often answers are given by a small amount of people (less then 5%). Many of these answers, but also some significant ones, are not (yet) explainable by the Weak Completion Semantics. A plausible explanation for that is that these people simply guess or use one of the heuristics mentioned below (educated guess).

A *generative approach* to model this lies in using MPTs. A MPT for a random guess can lead to all nine conclusions. MPTs for a particular heuristic strategy only take into account the valid conclusions under the corresponding theory. For the atmosphere bias, universal and affirmative conclusions are excluded when one of the premises is existential or negative, resp. In the case of identical moods, the conclusion must have this mood as well. For the matching bias, the following order from the most to the least conservative quantifier is defined on moods:

$$\mathsf{E} > \mathsf{O} = \mathsf{I} > \mathsf{A}$$

A conclusion may not be answered if it is *less conservative* than one of the premises wrt. that order. We have also observed biased conclusions in the data of [12] that may be explained by one of these heuristic strategies: in almost all syllogisms with figure 1, Xac is answered where X is the least conservative mood from the premises that is still allowed under the matching strategy (I is preferred over O). The answer Xca is not given at all.

As an alternative to generating the answers given by a cluster of guessers using MPTs, the following inversed process can be considered: predictions of the Weak Completion Semantics that are not in accordance with a particular heuristic strategy are not given by a cluster using that strategy. In the *filtering approach*, these conclusions are suppressed in the predictions for such a cluster. If no conclusion remains, NVC is answered instead. As it is likely that some participants does not use logic [20], such clusters must be modeled under the Weak Completion Semantics by using the generative of the filtering approach. As a consequence, MPTs can construct a prediction for all answer possibilities.

5.4 A Clustering Approach

Based on the principles and heuristic strategies described in this paper, the participants of [12] have been partitioned into three clusters using logic and two clusters applying heuristic strategies:

1. Basic principles, searchAlt, and converse for I
2. Basic principles, converse for I and deliberateGen
3. Basic principles, converse for I, E, and contraposition for A
4. Matching strategy
5. Biased conclusions in figure 1

Abduction was only used in one cluster because of the computational effort it requires. Although it would be interesting to model it for different clusters, except for converse, no other advanced principle would have an impact, because they do not add existential imports. According to the results of [1], abduction has the same results independent of whether only the converse I mood or both the converse I and E mood are used. The matching strategy was implemented using the filtering approach. The *biased conclusions in figure 1* heuristics was implemented using the generative approach such that its prediction overwrites the answers of other clusters, except NVC.

5.5 Evaluation

We evaluate the predictions of WCS based on the clustering approach described in Section 5.4. The prediction for the syllogism AO3 and the overall results are compared with other cognitive theories in Table 4. The Weak Completion Semantics predicts the participants' answers in [12] correctly for 33 out of the 64 Syllogisms. For 19 syllogisms there is one incorrect prediction, for 11 syllogisms there are two and for one syllogism there are three mismatches.

Syllogism	Participants	PSYCOP	Verbal Models	Mental Models	Conversion	WCS
AO3	Oca NVC	Oca Ica Iac	Oca NVC	Oca NVC Oac	Oca NVC	Oca NVC
Overall	100%	77 %	84 %	78 %	83%	92 %

Table 4. Comparison of the Weak Completion Semantics with other cognitive theories.

6 Conclusions

The starting point of this paper was the cognitive theory based on the Weak Completion Semantics and the principles defined in [1,2]. We have successfully extended this approach by introducing two new principles and applying a clustering approach to model individual differences in human reasoning. This also takes into account that some people may not use logic at all, but rather guess or apply heuristic strategies. The clustering presented in Section 5.4 is only the currently known best clustering under WCS but we don't know whether it is already the optimal one. However, due to the combinatorial explosion[1], it is difficult to find the global optimum. Future work may investigate alternative clusters and possibly identify new principles. The question whether the predictions change if abduction is applied to more than one cluster would be particularly interesting.

Finally, we have applied Multinomial Processing Trees to model that different principles lead to different conclusions. This information is lost if the predictions for all clusters are accumulated. This shows how much we depend on the way experimental results are reported. If we would have more insight about the patterns participants opted for, we could model single syllogisms by MPTs instead of fitting to the overall results.

References

1. Ana Costa, Emmanuelle-Anna Dietz, Steffen Hölldobler, and Marco Ragni. A computational logic approach to human syllogistic reasoning. In TBA, editor, *Proceedings of the 39th Annual Conference of the Cognitive Science Society, COGSCI 2017*, page TBA. Cognitive Science Society, 2017.
2. Ana Costa, Emmanuelle-Anna Dietz Saldanha, and Steffen Hölldobler. Monadic reasoning using weak completion semantics. In Steffen Hölldobler, Andrey Malikov, and Christoph Wernhard, editors, *Proceedings of the Young Scientist's Second International Workshop on Trends in Information Processing (YSIP2) 2017*. CEUR Workshop Proc., 2017.
3. 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, *Proc. of the 34th Annual Conference of the Cognitive Science Society*, pages 1500–1505, 2012.
4. Emmanuelle-Anna Dietz. A computational logic approach to the belief bias in human syllogistic reasoning. In Patrick Brézillon, Roy Turner, and Carlo Penco, editors, *10th International and Interdisciplinary Conference on Modeling and Using*

[1] For n principles, there are up to 2^n possible clusters. Additionally, it is unknown if the current set of principles is already complete.

Context CONTEXT 2017, Proceedings, volume 10257, pages 692–707. Springer, 2017.

5. Emmanuelle-Anna Dietz Saldanha, Steffen Hölldobler, and Luís Moniz Pereira. Contextual reasoning: Usually birds can abductively fly. In TBA, editor, *14th int. Conference on Logic Programming and Nonmonotonic Reasoning*, volume TBA of *LNAI*. Springer, 2017.
6. H. P. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and semantics*, volume 3. Academic Press, 1975.
7. S. Hölldobler. Weak completion semantics and its applications in human reasoning. In U. Furbach and Claudia Schon, editors, *Proceedings of the Workshop on Bridging the Gap between Human and Automated Reasoning on the 25th International Conference on Automated Deduction*, pages 2–16. CEUR-WS.org, 2015.
8. S. Hölldobler and C. D. Kencana Ramli. Logic programs under three-valued Łukasiewicz semantics. In P. M. Hill and D. S. Warren, editors, *25th int. Conference on Logic Programming,* LNCS 5649, Lecture Notes in Computer Science, pages 464–478. Springer, 2009.
9. S. Hölldobler and C. D. Kencana Ramli. Logics and networks for human reasoning. In *int. Conference on Artificial Neural Networks,* LNCS, 5769, pages 85–94. Springer, 2009.
10. Xiangen Hu and William H Batchelder. The statistical analysis of general processing tree models with the em algorithm. *Psychometrika*, 59(1):21–47, 1994.
11. P. N. Johnson-Laird. *Mental models: towards a cognitive science of language, inference, and consciousness.* Harvard University Press, Cambridge, MA, 1983.
12. S. Khemlani and P. N. Johnson-Laird. Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, pages 427–457, 2012.
13. J. W. Lloyd. *Foundations of Logic Programming.* Springer, 1984.
14. Luís M. Pereira, Joaquim Nunes Aparício, and José Júlio Alferes. Hypothetical reasoning with well founded semantics. In B. Mayoh, editor, *Scandinavian Conference on Artificial Intelligence: Proc. of the SCAI'91*, pages 289–300. IOS Press, Amsterdam, 1991.
15. Marco Ragni, Emmanuelle-Anna Dietz, Ilir Kola, and Steffen Hölldobler. Two-valued logic is not sufficient to model human reasoning, but three-valued logic is: A formal analysis. In Claudia Schon and Ulrich Furbach, editors, *Proceedings of the Workshop on Bridging the Gap between Human and Automated Reasoning co-located with 25th International Joint Conference on Artificial Intelligence IJCAI*, volume 1651 of *CEUR Workshop Proc.*, pages 61–73. CEUR-WS.org, 2016.
16. Marco Ragni, Henrik Singmann, and Eva-Maria Steinlein. Theory comparison for generalized quantifiers. In *CogSci*, 2014.
17. David M Riefer and William H Batchelder. Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95(3):318–339, 1988.
18. L. J. Rips. *The psychology of proof: Deductive reasoning in human thinking.* MIT Press, 1994.
19. K. Stenning and M. van Lambalgen. *Human Reasoning and Cognitive Science.* A Bradford Book. MIT Press, 2008.
20. N. E. Wetherick and K. J. Gilhooly. 'atmosphere', matching, and logic in syllogistic reasoning. *Current Psychology*, 14(3):169–178, 1995.
21. R. S. Woodworth and S. B. Sells. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 18(4):451, 1935.