TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Computer Science Institute of Artificial Intelligence Knowledge Representation and Reasoning

Clusters of Humans in Syllogistic Reasoning under the Weak Completion Semantics

Richard Mörbitz

KRR Report 17-02

Mail to
Technische Universität Dresden
01062 Dresden

Office
Room 2006
Nöthnitzer Straße 46
01187 Dresden

Internet
http://www.wv.inf.tu-dresden.de

Clusters of Humans in Syllogistic Reasoning under the Weak Completion Semantics

Richard Mörbitz

July 17, 2017

Contents

1 Introduction

The problem of understanding human reasoning processes is far from being solved. There are many cognitive theories, such as Mental Models [JL80] or PSYCOP [Rip94], that differ to a large degree. Unfortunately, there is no consensus on what theory is correct.

Syllogisms are a famous reasoning task that dates back to Aristotle. They have been studied by psychologists during the last two millennia and are well understood. In a recent meta-analysis, Khemlani and Johnson-Laird [KJL12] evaluated twelve cognitive theories. Seven of them have been compared with the answers of humans to syllogistic reasoning tasks. None of the tested theories was able to model human syllogistic reasoning adequately. It has been stated that a general theory of reasoning is of major importance to the Cognitive Science community.

The Weak Completion Semantics is a new cognitive theory that has its roots in a book on human reasoning by Stenning and van Lambalgen [SvL08]. The book had some technical mistakes, which were corrected by Hölldobler and Kencana Ramli [HR09] by using three-valued Łukasiewicz logic. Since then, the Weak Completion Semantics has been applied to many famous problems from Cognitive Science. These are, among others, the suppression task [DHR12], the selection task [DHR13], the belief-bias effect [PDH14a, PDH14b], reasoning about conditionals [DH15, DHP15], spatial reasoning [DHH15], and syllogistic reasoning [Die15, CDHR16]. The development of a general monadic reasoning theory based on the Weak Completion Semantics has been proposed recently [dCSH17].

The theory has assumed that all humans reason uniformly. There are, however, findings in psychology that support the thesis of individual differences between reasoners [JLS78, BHN03, KJL16]. This work introduces an approach to model these differences by grouping human reasoners into clusters. When looking at psychological studies on syllogisms, such as [WS35] and [WG95], it is reasonable to assume that not all humans use logic in reasoning task; some might apply heuristic strategies. In this work it is shown how such heuristics can be applied to the Weak Completion Semantics.

The representation of cognitive theories using formal logics is not always suitable to illustrate how certain conclusions are obtained in reasoning tasks and often it is difficult to understand for people without a background in formal logics. Therefore, the usage of multinomial processing trees to explain the outcomes of reasoning tasks has been suggested [RSS14]. How this can be done for the predictions of the Weak Completion Semantics is also a subject of this work.

The rest of the project report has the following structure: in Section 2, syllogisms are explained and the necessary background information from logic programming is given. In Section 3.1, it is explained how syllogistic reasoning tasks are solved under the Weak Completion Semantics. The concepts of human clusters and heuristic solving strategies will be introduced along with a proposal on how to model them under the Weak Completion Semantics. The representation of these processes as tree models will be illustrated in Section 4. The clustering approach is evaluated in Section 5 in terms of how well it predicts human reasoners. The report will be concluded in Section 6 with a summary and a discussion about future work.

2 Preliminaries

In this chapter the necessary background information is provided. From a psychological point of view, the syllogistic reasoning task is introduced and its importance for cognitive science is clarified. From formal logic, the concepts of logic programming, three-valued Łukasiewicz logic and the Weak Completion Semantics are explained. They are the foundation of our approach to model the syllogistic reasoning task.

2.1 Syllogistic Reasoning

Syllogisms are among the forms of reasoning that have been researched for the longest time, they have first been defined by Aristotle in ancient Greece. Since then, they have been investigated by both logicians [Łuk57] and psychologists [WS35, JLS78] in theoretical work and experimental studies. Thus syllogisms have become a central point in the attempt to formalize human reasoning.

2.1.1 Structure of Syllogisms

Definition 1 (Syllogism) *A syllogism is a logical argument that consists of three parts:*

- *major premise,*
- *minor premise,*
- *conclusion,*

each of which makes an assertion about two items.

Historically, the items were referred to as *major (predicate) term*, *minor (subject) term*, and *middle term*. The major premise links the predicate term with the middle term, the minor premise links the subject term with the middle term, and the conclusion is deduced knowledge about the subject term and the predicate term.

The syllogistic reasoning task is then formulated as follows: given the major premise and the minor premise, is the conclusion valid? As an example, consider the following classical syllogism:

All men are mortal.
All Greeks are men.
Therefore, all Greeks are mortal.

Here, *being mortal* is the major term, *Greeks* is the minor term, and *men* is the middle term. The premises do not contain a statement that directly connects the major and the minor term, but they are linked via the middle term. This enables the deduction of a valid conclusion.

Aristotle often used the Greek letters α, β, and γ as placeholders for terms instead of concrete items. In this connection, we draw on the Latin letters a, b, and c for abstraction, where a is the major term, b is the middle term and c is the minor term.

	Universal	Existential
Affirmative	All (A)	Some (I)
Negative	No (E)	Some not (O)

Table 1: Categorization of syllogisms by moods.

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

Table 2: Categorization of syllogisms by figures.

2.1.2 Categorizing Syllogisms

The premises and the conclusion can differ both in what quantification over the items is used and how the terms are arranged.

Classically there are four quantifiers that are called *moods*: *All* (A), *Some* (I), *No* (E), and *Some not* (O). They mirror the combinations of affirmative vs. negative and universal vs. existential quantification as presented in Table 1. Other quantifiers, such as *Few* and *Most*, are not considered here, although they are investigated in some studies [SMP94, KJL12]. Premises are abbreviated with their mood, e.g. 'All a are b' becomes Aab.

While premises were originally defined based on the term they contain (major vs. minor), another distinction can be made by the order of terms within such a premise. In each premise, the middle term can be in the first or second position, leading to four possible combinations. These combinations are called *figures* and assigned a number as presented in Table 2.

Consider the syllogism from above in its abstract form:

All b are a.
All c are b.
∴ All c are a.

The short notation is AA2, obtained by first listing the moods of the premises followed by the figure. Note that the categorization of syllogistic premises does not specify the conclusion.

2.1.3 Empirical Approach

A benefit of using syllogisms for investigating human reasoning is that while there is an infinite amount of possible syllogisms, there are only 512 distinct logical forms. This number is obtained by each possible combination of moods in the premises and the conclusion as well as the figure of the premises and the predicate and subject term of the conclusion ($4^3 \times 8 = 512$).

However, in many studies the participants are only given the two premises and asked which conclusions validly follow in their opinion. This reduces the number of logically distinct syllogistic reasoning tasks to 64. The set of possible answers consists of eight conclusions (four moods and the possibility to swap the predicate and the subject term) and *no valid conclusion* (NVC), representing the opinion that all other conclusions are invalid.

The large amount of psychological studies on syllogisms allows to retrieve representative data on how humans solve this task. Many of those studies have been accumulated in a meta-analysis by Khemlani and Johnson-Laird in 2012 [KJL12], who seek to develop a unified theory of reasoning [JLK13]. Syllogisms are suitable for this task, since they are well understood (recall more than 2000 years of research) while complex enough to be a serious cognitive task.

Khemlani and Johnson-Laird came to the conclusion that none of the twelve investigated cognitive theories could correctly predict how humans reason about syllogisms. They close with a call for a better, comprehensive, and unified theory of human reasoning. Any new theory, such as our approach with the Weak Completion Semantics, will have to prove its value by predicting the answers of the participants better than the existing theories. How the fit of a theory to that data is computed will be described in Section 5.

2.2 Weak Completion Semantics

This section introduces the formal concepts that are necessary to understand our approach to human reasoning. Starting with logic programs, the ideas of using a three-valued logic and 'weakly completing' a program will be explained. Based on that, it will be shown how the two reasoning forms *deduction* and *abduction* are modelled under the Weak Completion Semantics.

2.2.1 Logic Programs

The reader is expected to be familiar with the basic notions from first-order logic, namely *constant symbol*, *variable*, *predicate*, *atom*, *literal*, *clause*, *head*, *body*, *formula*, and *quantifier*. Understanding for the semantics of the truth-value constants $\top$ and $\bot$ as well as the logic connectives $\neg$, $\wedge$, $\vee$, $\leftarrow$, and $\leftrightarrow$ is needed, too. For a broad introduction to classical logic, see e.g. [Llo87, Höl09].

Definition 2 (Logic Program) *A logic program $\mathcal{P}$ is a finite set of clauses. Each clause is of one of the following forms:*

1. $A \leftarrow \top$ (Fact),
2. $A \leftarrow \bot$ (Assumption),
3. $A \leftarrow B_1 \wedge \cdots \wedge B_n$, (Rule),

where $n > 0$, A is an atom, and B_i are literals for $1 \leq i \leq n$.

Facts are objective knowledge, the atom in the head of a fact is equivalent to true under the Weak Completion Semantics. Assumptions may seem like negative facts, however, the atom in the head is only equivalent to false if there is no other clause in the program that has the same atom in its head.[1] Rules allow the inference of new knowledge from facts and assumptions.

Only datalog programs are considered, so only constant symbols and universally quantified variables, but no function symbols are allowed in terms.

For reasoning, instead of working with the logic program directly, its corresponding *ground program* is used.

Definition 3 (Ground Program) *A ground program $g\mathcal{P}$ is the set of all ground instances of the clauses occurring in the logic program $\mathcal{P}$.*

A ground instance of a clause C is obtained by replacing all variables occurring in C with constant symbols. Since the set of constant symbols if finite, the ground program $g\mathcal{P}$ of a program $\mathcal{P}$ is finite as well.

In a ground program, two particular sets of ground atoms are of special interest: *defined* and *undefined* atoms.

Definition 4 (Defined Atom) *Let $\mathcal{P}$ be a logic program. An atom A is* defined *in $g\mathcal{P}$ if and only if $g\mathcal{P}$ contains a clause with A in its head that is a fact or a rule.*

Definition 5 (Undefined Atom) *Let $\mathcal{P}$ be a logic program. An atom A is* undefined *in $g\mathcal{P}$ if and only if it is not defined in $g\mathcal{P}$.*

For defined atoms, it is sometimes interesting to know the clauses that are responsible for its definition.

Definition 6 (Definition of an Atom) *Let $\mathcal{P}$ be a logic program and A be an atom. The* definition of A in $\mathcal{P}$ *is the following set:*

$$\mathrm{def}(A, \mathcal{P}) = \{A \leftarrow body \mid A \leftarrow body \textit{ is a rule or a fact in } g\mathcal{P}\}$$

Based on these definitions, the concept of assumptions can now be formally defined.

Definition 7 (Assumed Literal) *Let $\mathcal{P}$ be a logic program and A be an atom. $\neg A$ is* assumed *in $\mathcal{P}$ if and only if A is undefined in $g\mathcal{P}$ and $g\mathcal{P}$ contains an assumption with A in its head, i.e., $def(\mathcal{A}, \mathcal{P}) = \emptyset$ and $A \leftarrow \bot \in g\mathcal{P}$.*

[1] This capability of overwriting assumptions is obtained by using a completion semantics.

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

$\wedge$	$\top$	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	$\top$

Table 3: The truth tables for the logic connectives under Ł-logic.

2.2.2 Three-Valued Logic and Models

Logic programs are used with a special three-valued logic defined by Łukasiewicz [Łuk20] (Ł-logic). It contains the following truth values: *true* ($\top$), *false* ($\bot$), and *unknown* (U). Like in classical logic, formulae are interpreted to obtain their truth value.

Definition 8 (Three-Valued Interpretation) *A three-valued interpretation I under Ł-logic is a mapping from the set of formulae to the set $\{\top, \bot, \mathsf{U}\}$. The truth value of a formula F under I is obtained by evaluating the logic connectives occurring in F as defined in Table 3.*

Three-valued interpretations are represented as tuples. Let F be a formula, its interpretation is $I = \langle I^\top, I^\bot \rangle$, where

$$I^\top = \{A \mid A \text{ is an atom occurring in } F \wedge I(A) = \top\},$$

$$I^\bot = \{A \mid A \text{ is an atom occurring in } F \wedge I(A) = \bot\},$$

$$I^\top \cap I^\bot = \emptyset.$$

Any atom that does not occur in $I^\top \cup I^\bot$ is implicitly mapped to U.

Logic programs are considered under a model semantics.

Definition 9 (Model) *Let $\mathcal{P}$ be a logic program and I be a three-valued interpretation. I is a* model *of $\mathcal{P}$ if and only if $I(C) = \top$ for every clause C in $g\mathcal{P}$.*

Towards Least Models Minimality properties of models are of particular interest. Therefore, a partial order of interpretations is defined. Let I, J be three-valued interpretations. $I \subseteq J$ if and only if $I^\top \subseteq J^\top$ and $I^\bot \subseteq J^\bot$.

Finally, the concept of a *least* model that will be used for reasoning can be defined.

Definition 10 (Least Model) *Let $\mathcal{P}$ be a logic program and I be a three-valued interpretation, such that I is a model of $\mathcal{P}$. I is the* least model *of $\mathcal{P}$ if and only if for every other interpretation J that is a model of $\mathcal{P}$, $I \subseteq J$ holds.*

2.2.3 Reasoning with Respect to Least Models

For modelling human reasoning processes, instead of the least model of a logic program, the least model of its *weak completion* is considered.

Definition 11 (Completion of a Program) *Let $\mathcal{P}$ be a logic program. The* comple-tion *of $\mathcal{P}$, denoted by $c\mathcal{P}$, is obtained from $\mathcal{P}$ by applying the following steps:*

1. *For each atom A in $g\mathcal{P}$, replace all clauses of the form $A \leftarrow Body$ in $g\mathcal{P}$ by the clause $A \leftarrow \bigvee_{A \leftarrow Body \in g\mathcal{P}} Body$.*

2. *For all atoms A that are not defined in $g\mathcal{P}$, add a clause $A \leftarrow \bot$.*

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

This completion dates back to Clark [Cla78]. The reasoning behind it is that implications are implicitly meant to be equivalences, but the inverse conditional is omitted by the author of the logic program. Additionally, it corresponds to the concept of the closed world assumption, which allows to assume that any undefined atom in a logic program is *false*.

That is, however, inadequate for modelling human reasoning. Atoms for which no knowledge may be derived in the program should instead be regarded as *unknown*. This has been done by Hölldobler and Kencana Ramli [HR09] by introducing the concept of the *weak completion* ($wc\mathcal{P}$) of a program. It is obtained just as the completion, but step 2 is omitted.

Least models of the weak completion of a program are used for *forward reasoning* (deduction), one of the three forms of reasoning identified by Peirce [Pei74]. The other forms are *backward reasoning* (abduction), which is introduced in Section 2.2.4, and induction, which is not considered in this work.

It has been shown [HR09] that a least model always exists for the weak completion of a program. This least model coincides with the least fixed point of the immediate consequence operator defined by Stenning and van Lambalgen [SvL08].

Definition 12 (Immediate Consequence Operator ($\Phi_\mathcal{P}$)) *Let $\mathcal{P}$ be a logic program and I be a three-valued interpretation. Then, $\Phi_\mathcal{P}(I) = \langle J^\top, J^\bot \rangle$, where*

$$
\begin{aligned}
J^\top &= \{A \mid \textit{there exists a clause } A \leftarrow \textit{body in } g\mathcal{P} \textit{ such that } I(\textit{body}) = \top\},\\
J^\bot &= \{A \mid \textit{there exists a clause } A \leftarrow \textit{body in } g\mathcal{P} \textit{ and}\\
&\qquad\qquad \textit{for all clauses } A \leftarrow \textit{body in } g\mathcal{P} \textit{ it holds that } I(\textit{body}) = \bot\}.
\end{aligned}
$$

For finite datalog programs, as they are used in this work, the least fixed point of $\Phi_\mathcal{P}$ always exists. The least model computed by $\Phi_\mathcal{P}$ when starting with the empty interpretation $\langle \emptyset, \emptyset \rangle$, denoted by $\mathcal{M}_\mathcal{P}$, is the minimal knowledge that is inferred from the program $\mathcal{P}$. $\mathcal{M}_\mathcal{P}$ is the result of deduction. All formulae that are *true* under $\mathcal{M}_\mathcal{P}$ are said to be *entailed* by $\mathcal{P}$ under the Weak Completion Semantics.

Definition 13 (Entailment Relation $\models_{wcs}$) *Let $\mathcal{P}$ be a logic program, F be a formula, and $\mathcal{M}_\mathcal{P}$ be a three-valued interpretation, such that $\mathcal{M}_\mathcal{P}$ is the least model of $wc\mathcal{P}$. Then, $\mathcal{P} \models_{wcs} F$ if and only if $\mathcal{M}_\mathcal{P}(F) = \top$.*

2.2.4 Reasoning with Abduction

Backward reasoning (abduction) is the process of deriving new knowledge that is not guaranteed by the premises. Given a logic program and an observation that does not follow from the program, an explanation for that observation is searched. If one is found that is consistent with the original program, it may be added to the knowledge base.

The process of abduction is formalized as an *abductive framework*.

Definition 14 (Abductive Framework) *An abductive framework is a quadruple of the form* $\langle \mathcal{P}, \mathcal{A}, \mathcal{IC}, \models \rangle$*, where*

> $\mathcal{P}$ *is logic program,*
>
> $\mathcal{A}$ *is a finite set of formulae called* abducibles*,*
>
> $\mathcal{IC}$ *is a set of* integrity constraints *(see Definition 16),*
>
> $\models$ *is a logical entailment relation.*

Although any set of formulae could serve as abducibles, only a particular set is considered: the *abducibles with respect to* $\mathcal{P}$.

Definition 15 (Abducibles) *Let* $\mathcal{P}$ *be a logic program. The set of abducibles with respect to* $\mathcal{P}$*, denoted by* $\mathcal{A}_\mathcal{P}$*, is defined as follows:*

$$\begin{aligned} \mathcal{A}_\mathcal{P} = {} & \{A \leftarrow \top \mid A \textit{ is undefined in } \mathcal{P}\} \cup \\ & \{A \leftarrow \bot \mid A \textit{ is undefined in } \mathcal{P}\} \cup \\ & \{A \leftarrow \top \mid \neg A \textit{ is assumed in } \mathcal{P}\} \end{aligned}$$

Integrity constraints are special expressions that can be use to add further restrictions to the abduction process.

Definition 16 (Integrity Constraint) *An integrity constraint is a clause of the form* $\mathsf{U} \leftarrow B_1 \wedge \cdots \wedge B_n$*, where* $n > 0$*,* B_i *is a literal for all* $1 \leq i \leq n$*, and* U *is the truth-value constant denoting the unknown.*

An interpretation I *satisfies* a set of integrity constraints $\mathcal{IC}$ if and only if $I(C) = \top$ for all clauses $C \in \mathcal{IC}$. Note that if all of the literals in the body of an integrity constraint are mapped to *true*, the truth value of the integrity constraint is *unknown* and I does not satisfy $\mathcal{IC}$. However, it is possible that literals in the body are *unknown*, because $\mathsf{U} \leftarrow \mathsf{U}$ is evaluated to *true* under Ł-logic.

With the definitions from above, the abductive framework used under the Weak Completion Semantics is instantiated as $\langle \mathcal{P}, \mathcal{A}_\mathcal{P}, \emptyset, \models_{wcs} \rangle$, given a logic program $\mathcal{P}$. While any set of literals could theoretically be used as an observation, we restrict them to certain sets of atoms that are obtained from $\mathcal{P}$. We define the set of observations in the following way:

Definition 17 (Observations) *Let $\mathcal{P}$ be a logic program. The set of observations with respect to $\mathcal{P}$, denoted $\mathcal{O}_\mathcal{P}$, is defined 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 $1 \leq i \leq n$.

Intuitively, these are the atoms that occur in the head of both a rule and a fact. The set of observations is further restricted by considering only facts that result from certain principles. See Section 3.2.3 for an example.

Given such an observation $\mathcal{O} \in \mathcal{O}_\mathcal{P}$, the task of the abductive framework is to find an explanation for it that meets certain requirements.

Definition 18 (Explanation) *Let $\langle \mathcal{P}, \mathcal{A}_\mathcal{P}, \mathcal{IC}, \models_{wcs} \rangle$ be an abductive framework and $\mathcal{O}$ be a literal (observation). $\mathcal{O}$ is* explainable *in the abductive framework $\langle \mathcal{P}, \mathcal{A}_\mathcal{P}, \mathcal{IC}, \models_{wcs} \rangle$ if and only if there exists an $\mathcal{E} \subseteq \mathcal{A}$, such that:*

1. $\mathcal{P} \cup \mathcal{E} \models_{wcs} \mathcal{O}$,
2. $\mathcal{P} \cup \mathcal{E}$ *satisfies* $\mathcal{IC}$.

$\mathcal{E}$ is then called explanation *for $\mathcal{O}$.*

Since there may be several explanations for an observation, a guideline for drawing conclusions from explanations is introduced. First of all, it is assumed that humans prefer minimal explanations for reasoning.

Definition 19 (Minimal Explanation) *Let $\langle \mathcal{P}, \mathcal{A}_\mathcal{P}, \mathcal{IC}, \models_{wcs} \rangle$ be an abductive framework, $\mathcal{O}$ be a literal (observation), and $\mathcal{E} \subseteq \mathcal{A}$ be an explanation for $\mathcal{O}$. $\mathcal{E}$ is* minimal *if and only if there exists no 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 given $\mathcal{P}$, $\mathcal{O}$, and $\mathcal{IC}$. F follows *skeptically*, if it is entailed by all explanations given $\mathcal{P}$, $\mathcal{O}$, and $\mathcal{IC}$.

Due to the results of [dCSH17], skeptical abduction is used.

3 Human Reasoning Processes

In the following the approach of modelling human reasoning under the Weak Completion Semantics will be explained. Syllogisms are understood as *monadic quantified assertions*. This means that they are formalized as universally quantified logic clauses. All predicate symbols in the logic program have an arity of one and only constant symbols and variables, but no function symbols are used in terms.

In the beginning, an overview of already known principles in human reasoning is given. After that, it is shown how clusters of humans can be modelled with the help of such principles and two new principles are introduced. In the end, heuristic strategies that do not involve logic, but can still be used to solve reasoning tasks, are presented.

3.1 Common Principles of Human Reasoning

Eight principles of reasoning have already been identified from findings in Cognitive Science and Psychology [dCSH17]. They are introduced in this section.

Definition 20 (Principle of Reasoning) *A principle is a modular component of a reasoning process that is represented as a set of clauses.*

Several principles can be combined with each other to model how a reasoner solves a reasoning task. The union of the sets of clauses representing each of it is the logic program that encodes the reasoning task.

Based on this logic program, the reasoning forms *deduction* and *abduction* can be simulated. Note that in [dCSH17] an additional logic program was used to obtain the results of deduction based on the least model of the first logic program. The investigation of the principles used for that additional logic program and how they can be combined, however, is beyond the scope of this project. Therefore, that way of deduction is not considered here; instead, the conclusions are entailed from the least model directly.

For an overview of principles and their representation as clauses, see Table 4. In the following, the motivation behind each principle is described. After that, an example of encoding a syllogism and applying deduction and abduction to obtain the conclusions is given.

3.1.1 Quantified Assertion as Conditional

A quantified assertion, e.g. “All a are b”, contains statements about two predicates, namely a and b. In a logic program clause, only one of them can be in the head and the other one must be in the body, thus forming a conditional.

The representation of such a conditional is as follows: if a quantified assertion establishes a relation about the terms y and z, the first term y is seen as the antecedent and the second term z is seen as the conclusion of a conditional. The formalization as a monadic quantified conditional is then: $z(X) \leftarrow y(X)$. Intuitively, if we know that an object X belongs to the term y, we also deduce that it belongs to z.

Note that in this encoding, it is impossible to determine what syllogistic mood the assertion had. This is solved by the following principles.

Principle of Reasoning	Corresponding Clauses
Quantified assertion as conditional + licenses	$z(X) \leftarrow y(X) \wedge \neg ab_{yz}(X)$
Existential import + licenses	$ab_{yz}(o_1) \leftarrow \bot$ $y(o_1) \leftarrow \top$
Unknown generalization + licenses	$y(o_2) \leftarrow \top$
No refutation + licenses	$ab_{yz}(X) \leftarrow \bot$
Negative quantified assertion + licences	$z'(X) \leftarrow y(X) \wedge \neg ab_{nyz}(X)$
Negation by transformation + licenses	$z(X) \leftarrow \neg z'(X) \wedge \neg ab_{nzz}(X)$
No derivation by double negation + licenses	$ab_{nzz}(o_{1/2}) \leftarrow \bot$

Table 4: Reasoning principles and their representation as sets of clauses.

3.1.2 Licenses for Inferences

Stenning and van Lambalgen [SvL08] proposed to see conditionals as *licenses for inferences*. A monadic quantified conditional "For all X, $z(X)$ holds if $y(X)$ holds" is replaced by "For all X, $z(X)$ holds if $y(X)$ holds and nothing is abnormal with X". This is formalized by introducing an abnormality predicate $ab_{yz}(X)$. The license is then implemented as a conjunction in the body of the conditional: $z(X) \leftarrow y(X) \wedge \neg ab_{yz}(X)$. The abnormality plays an important role, because it has to be *false* to enable any inference about $z(X)$. This is achieved by the *negative assumptions* introduced by the following principles. For such clauses, we will write that they origin from both the *licenses by inferences* and the corresponding other principle.

3.1.3 Existential Import

In classical logic, a universally quantified formula is also valid if the set of objects over which is quantified is empty. Humans do not seem to follow this logic, because a quantification over some things in natural language is done with the intention that these things exist. This phenomenon is called Gricean Implicature [Gri75], because a universal quantifier seems to imply an existential quantifier.

In the previous principles, conditionals of the form $z(X) \leftarrow y(X)$, that are regarded as universally quantified, have been introduced. The required existential import is encoded by the fact $y(o) \leftarrow \top$, where o is a new object that does not yet appear in the logic program. If licenses are used, it would still be impossible to infer $z(o)$ in such a conditional, because the abnormality of o with respect to y and z is unknown. However, since the original assertion states that a certain relation of y and z holds, we assume that this abnormality does not hold for the imported object o. This is formalized by the assumption $ab_{yz}(o) \leftarrow \bot$.

3.1.4 Unknown Generalization

There is a logical difference between "all y are z" and "some y are z" that is also observable in the way humans answer syllogistic reasoning tasks [KJL12]. However, if only the principles *quantified assertion as conditional*, *licenses for inferences*, and *existential import* were used, both cases would have the same encoding and thus the same conclusions.

This is solved by importing another object that is different from the already existing one. Formally, if o_1 was introduced by the *existential import* principle, a fact $y(o_2) \leftarrow \top$ is added if the mood is existentially quantified. Since nothing is stated about the abnormality, $z(o_2)$ will remain *unknown*. As a consequence, there is an object in y which is known to be in z and one for which this is unknown. Under three-valued logic, this fits to the existentially quantified mood.

3.1.5 No Refutation

According to the Mental Models theory, counterexamples are used spontaneously for reasoning [JL80]. This is called refutation by counterexample and covered by the use of licenses. As a consequence, any object can be used as a counterexample, except the one introduced by the *existential import* (because no abnormality is assumed for it).

However, this contradicts with universally quantified moods, as we do not want any object to be used as a counterexample for conditional such as $z(X) \leftarrow y(X) \wedge \neg ab_{yz}(X)$ if y and z are in a relation like 'all'. Therefore, the assumption $ab_{yz}(X) \leftarrow \bot$ is added for universally quantified moods.

3.1.6 Negation by Transformation

Logic program clauses may not have negated atoms in the head. This makes it impossible to encode premises with negative moods, because the negated atom would have to be in the head. Consider e.g. the premise "no y are z", whose corresponding conditional with licenses is $\neg z(X) \leftarrow y(X) \wedge \neg ab_{yz}(X)$.

To circumvent this, for each negative literal an additional atom is introduced, which instead is placed in the head of the clause. Conditionals like the one from above are then formulated as $z'(X) \leftarrow y(X) \wedge \neg ab_{yz}(X)$.

In order to be able to infer something about $z(X)$, an additional clause $z(X) \leftarrow \neg z'(X) \wedge \neg ab_{nzz}(X)$ is added. Now, if $z'(X)$ is inferred, the body of this clause is evaluated to *false*. Under the Weak Completion Semantics, $z(X)$ also becomes *false* if there is no other rule with it in the head.

Note that this principle has a technical origin and is only used to allow negative inferences. The only backup from Cognitive Science is that humans indeed draw conclusions in syllogisms with negative premises [KJL12].

3.1.7 No Derivation By Double Negation

Under the Weak Completion Semantics, derivation through double negation is possible. From two assertions like "if not a, then b" and "if not b, then c" it is possible to conclude that c is *true*, given that a is *true*.

The data in the meta-analysis on syllogistic reasoning [KJL12] shows that humans do not seem to infer knowledge through double negation. This is accounted by using licenses for all conditionals. In negative moods, where the *negation by transformation* principle must be used, the additional abnormalities $ab_{nzz}(X)$ are only assumed to be *false* for the imported objects and for no other X. Therefore, the atom $z(X)$ in the head of the rule introduced by that principle cannot become *true*.

3.1.8 Converse Interpretation

The premises "some y are z" and "some z are y" are logically equivalent. The same holds for premises with the mood E ('no'). If humans also reason that way, the converse premises Izy or Ezy, respectively, must be encoded and added to the logic program.

Both cases (adding the converse interpretation for I or E) can be considered independently from each other. There may also be syllogisms for which the converse interpretation is considered and others for which it is not. Finally, while not logically correct, humans could use the converse of premises with A and O moods for reasoning as well.

There is evidence that humans apply this principle in the experimental data [KJL12] to solve some syllogisms, because if the principle is applied to premises with the moods I and E, a better fit is achieved (see Section 5).

3.2 Entailment of Conclusions

It has all ready been said that the representations of all principles used to solve a syllogism are united to form a logic program. There are several possibilities to draw conclusions from the least model of such a program. In this work, an entailment as defined in first-order logic is considered.

The possible conclusions are drawn if and only if the corresponding formula is evaluated to *true* (where (y, z) is instantiated as (a, c) or (c, a)).

Ayz $\exists X(\mathcal{P} \models_{wcs} y(X)) \wedge \forall X(\mathcal{P} \models_{wcs} y(X) \rightarrow \mathcal{P} \models_{wcs} z(X))$

Eyz $\exists X(\mathcal{P} \models_{wcs} y(X)) \wedge \forall X(\mathcal{P} \models_{wcs} y(X) \rightarrow \mathcal{P} \models_{wcs} \neg z(X))$

Iyz $\exists X(\mathcal{P} \models_{wcs} y(X) \wedge z(X)) \wedge \exists X(\mathcal{P} \models_{wcs} y(X) \wedge \mathcal{P} \not\models_{wcs} z(X)) \wedge$
$\exists X(\mathcal{P} \models_{wcs} z(X) \wedge \mathcal{P} \not\models_{wcs} y(X))$[2]

Oyz $\exists X(\mathcal{P} \models_{wcs} y(X) \wedge \neg z(X)) \wedge \exists X(\mathcal{P} \models_{wcs} y(X) \wedge \mathcal{P} \not\models_{wcs} \neg z(X))$

NVC None of the above conclusions is entailed

3.2.1 AO3 — no valid conclusion

The syllogism AO3 consists of two premises:

All a are b.
Some c are not b.

[2]This third part of the conjunction is only used if the principle *converse interpretation* is applied.

Each of the premises is encoded according to the principles of reasoning that are applied. The first premise is represented as the following logic program, where for each clause the principle which added it is written:

$$
\begin{array}{rlr}
b(X) & \leftarrow a(X) \wedge \neg ab_{ab}(X) & \textit{conditional + licenses} \\
a(o_1) & \leftarrow \top & \textit{existential import} \\
ab_{ab}(o_1) & \leftarrow \bot & \textit{existential import + licenses} \\
ab_{ab}(X) & \leftarrow \bot & \textit{no refutation + licenses}
\end{array}
$$

The *quantified assertion as conditional* and *existential import* principles must be used for every syllogism to enable inference. The placeholders y and z are replaced by a and b, respectively, the terms of the premise. The objects are assumed to be named o_i, where i is incremented by one with each principle introducing a new object. Since the mood A is universally quantified, the *no refutation* principle is applied as well.

The second premise is encoded by:

$$
\begin{array}{rlr}
b'(X) & \leftarrow c(X) \wedge \neg ab_{cnb}(X) & \textit{negative conditional + licenses} \\
ab_{cnb}(o_2) & \leftarrow \bot & \textit{existential import + licenses} \\
c(o_2) & \leftarrow \top & \textit{existential import} \\
c(o_3) & \leftarrow \top & \textit{unknown generalization} \\
b(X) & \leftarrow \neg b'(X) \wedge \neg ab_{nbb}(X) & \textit{negation by transformation + licenses} \\
ab_{nbb}(o_2) & \leftarrow \bot & \textit{no derivation by double negation + licenses} \\
ab_{nbb}(o_3) & \leftarrow \bot & \textit{no derivation by double negation + licenses}
\end{array}
$$

Since the mood is is negative, the *quantified assertion as conditional* principle is used with the alternative atom in the head of the clause. Again, the placeholders are replaced by the terms of the premise. As for all syllogisms, the *existential import* principle is used, but with o_2 as its object (the enumeration is continued from the previous premise).

As the mood is existential, the *unknown generalization* principle is applied as well introducing a new object o_3. The negative mood is accounted with the *negation by transformation* principle. Finally, its abnormalities are assumed to be *false* for all objects introduced by this premise, but no others. This is the result of the *no derivation by double negation.*

The logic program representing the syllogism AO3, $\mathcal{P}_{AO3}$, is the union of the sets of clauses for each premise. The least model of the weak completion of $\mathcal{P}_{AO3}$ is:

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

Possible conclusions from this model are based on the atoms highlighted in gray. It can be seen that no valid conclusion (NVC) follows, because there is no o_i for which anything about $a(o_i)$ and $c(o_i)$ is known at the same time. The significant answers by the participants of the meta-analysis by Khemlani and Johnson-Laird [KJL12] are NVC and Oca. However, even if abduction was applied, Oca could not be concluded. This problem will be solved during this work.

3.2.2 AO4 — conclusion by deduction

AO4 is a syllogism which is similar to AO3:

All b are a.
Some b are not c.

Its encoding $\mathcal{P}_{AO4}$ is exactly as for AO3, only that the symbols representing the two terms are swapped in both premises. The weak completion of $\mathcal{P}_{AO4}$ has the following least model:

$$\langle\{\boxed{a(o_1)}, \boxed{a(o_2)}, a(o_3), b(o_1), b(o_2), b(o_3), c'(o_2)\}$$
$$\{ab_{ba}(o_1), ab_{ba}(o_2), ab_{ba}(o_3), ab_{bnc}(o_2), ab_{ncc}(o_2), ab_{ncc}(o_3), \boxed{c(o_2)}\}\rangle$$

Let this model be denoted $\langle I^\top, I^\bot\rangle$ for this paragraph. Consider the atoms highlighted in gray: $a(o_1) \in I^\top$, $a(o_2) \in I^\top$, $c(o_2) \in I^\bot$, but $c(o_1) \notin I^\bot$, so the conclusion Oac ("some are not c") can be drawn. This corresponds to the only significant answer by the participants.

3.2.3 AO1 — conclusion by abduction

For an example with abduction, consider the syllogism AO1:

All a are b.
Some b are not c.

Again, the encoding $\mathcal{P}_{AO1}$ is similar to $\mathcal{P}_{AO3}$, just swap the symbols representing the terms in the second premise. The least model of the weak completion of $\mathcal{P}_{AO1}$, denoted $\mathcal{M}_{\mathcal{P}_{AO1}}$, is as follows:

$$\langle\{\boxed{a(o_1)}, b(o_1), b(o_2), b(o_3), c'(o_2)\},$$
$$\{ab_{ab}(o_1), ab_{ab}(o_2), ab_{ab}(o_3), ab_{bnc}(o_2), ab_{ncc}(o_2), ab_{ncc}(o_3), \boxed{c(o_2)}\}\rangle$$

Similar to the syllogism AO3, no valid conclusion follows from this model. As a consequence, abduction is applied to search for alternatives.

The abductive framework is instantiated as $\langle\mathcal{P}_{AO1}, \mathcal{A}_{\mathcal{P}_{AO1}}, \emptyset, \models_{wcs}\rangle$, where

$$\begin{aligned}\mathcal{A}_{\mathcal{P}_{AO1}} = \; & \{ab_{ab}(oi) \leftarrow \top \mid i \in \{1,2,3\}\} \cup \\ & \{ab_{bnc}(o2) \leftarrow \top\} \cup \\ & \{ab_{ncc}(oi) \leftarrow \top \mid i \in \{1,2\}\} \cup \\ & \{ab_{bnc}(oi) \leftarrow \top, ab_{bnc}(oi) \leftarrow \bot \mid i \in \{1,3\}\} \cup \\ & \{ab_{ncc}(o1) \leftarrow \top, ab_{ncc}(o1) \leftarrow \bot\} \cup \\ & \{a(oi) \leftarrow \top, a(oi) \leftarrow \bot \mid i \in \{2,3\}\}\end{aligned}$$

The only possible observation is $b(o_2)$, as it is both head of a fact introduced by the *existential import* principle and head of a clause that is not a fact. The only explanation for $b(o_2)$ is $a(o_2)$, which trivially is also a minimal explanation. Therefore, $a(o_2)$ follows both skeptically and credulously and is added to the least model. The updated least model $\langle I^\top, I^\perp \rangle = \mathcal{M}_{\mathcal{P}_{AO1}} \cup \{a(o_2)\}$ reads as follows:

$$\langle \{ \boxed{a(o_1)}, \boxed{a(o_2)}, b(o_1), b(o_2), b(o_3), c'(o_2)\},$$
$$\{ab_{ab}(o_1), ab_{ab}(o_2), ab_{ab}(o_3), ab_{bnc}(o_2), ab_{ncc}(o_2), ab_{ncc}(o_3), \boxed{c(o_2)} \}\rangle$$

Again, the atoms important for deduction are highlighted in gray. Since $a(o_1) \in I^\top$, $a(o_2) \in I^\top$, $c(o_2) \in I^\perp$, but $c(o_1) \notin I^\perp$, the conclusion Oac is drawn instead. This is indeed the only significant answer of the participants.

3.3 Clusters of Reasoners

Differences between individuals in reasoning have been investigated in psychology for several times, consider [JLS78, BHN03, KJL16]. Due to differences in training in logic, motivation or for various other reasons people behave differently when they solve a syllogistic reasoning task and as a result, come do different conclusions.

When looking at the results of [KJL12], it can be observed that in 37 out of the 64 syllogisms NVC was a significant answer as well as another conclusion. As an example, see the syllogism AO3 introduced above: 40 % of the participants answer Oca (Some c are not a), while 20 % give NVC as an answer. Since these two answers contradict with each other it must be assumed that there were at least two groups of reasoners. One of them had the ability to draw a conclusion (Oca) that the other could not (thus answering NVC). In addition to that, in many psychological studies on syllogisms people are only allowed to give one answer. If all humans would reason in the same way, there would only be one significant answer to each syllogism. However, in 48 out of the 64 syllogisms, at least two answers are given.

It is proposed to account for that in the Weak Completion Semantics by using the existing concept of principles of reasoning to form *clusters of reasoners*.

Definition 21 (Cluster) *A cluster of human reasoners is a group of people that uses the same principles of reasoning to solve a task.*

The goal is to define a cluster for each answer of a syllogism that is given by a significant amount of participants. The cluster is given as a set of principles that entails the conclusion under the Weak Completion Semantics.

Until now, all principles have been applied in general, yielding one single model and thus one answer (possibly containing more than one conclusion) for each syllogism. With the introduction of clusters, this will no longer be the case. Since the principles vary between clusters, each cluster of reasoners will now have its own logic program representing the principles used. As a consequence, for each cluster the least model will be computed. The resulting individual sets of answers possibly differ. Just as all the individual answers are accumulated in psychological studies, the answers predicted for each cluster will be united to form a general prediction under the Weak Completion Semantics.

The principles *conditional by licenses*, *existential import*, *unknown generalization*, *negation by transformation* and *no derivation through double negation* are considered as *basic principles* that still need to be applied in every case, because they form the core of the logical inference. The *converse interpretation* principle is seen as one whose application may vary between clusters.

This section introduces two new principles to illustrate how the approach can be adapted to model clusters of reasoners.

3.3.1 The Context Operator

The `context` operator is a truth-functional operator extending three-valued logic programs that was introduced by Dietz, Hölldobler and Pereira [DHP17]. It is defined as follows:

$$\mathsf{ctxt}(L) = \begin{cases} \top & I(L) = \top, \\ \bot & \textit{otherwise}. \end{cases}$$

with respect to an interpretation I, where L is a literal.

Definition 22 (Contextual logic program) *A contextual logic program is a logic program where the* context *operator can occur in the body of clauses, i.e. additionally to the original clauses, clauses of the following form are also valid:*

$$A \leftarrow L_1 \wedge \cdots \wedge L_m \wedge \mathsf{ctxt}(L_{m+1}) \wedge \cdots \wedge \mathsf{ctxt}(L_{m+n}),$$

where $m \geq 0$, $n \geq 0$, $m + n > 1$, A is an atom and B_i are literals for $1 \leq i \leq m + n$.

We also have to extend the concept of assumptions. Until now, we called clause of the form $A \leftarrow \bot$, where A is an atom, assumptions. These are actually *negative assumptions*, because their intended meaning is to assume $\neg A$. In contrast to that we introduce clauses of the form $A \leftarrow \mathsf{U}$, where A is an atom, as *unknown assumptions*. They are needed when we want to block that the *context* operator adds negative assumptions to the program. Consider the following example:

$$\mathcal{P}_{ctxt} = \{ab(X) \leftarrow \mathsf{ctxt}(c(X))\}$$

Since nothing about $c(X)$ is known, $ab(X)$ will be *false* for all X. Imagine we do not want that predicate to be *false* for a specific object o. We then add the unknown assumption $ab(X) \leftarrow \mathsf{U}$ to $\mathcal{P}_{ctxt}$ and obtain $\mathcal{P}'_{ctxt}$. The weak completion of $\mathcal{P}'_{ctxt}$ is as follows (assuming o is the only constant symbol for simplicity):

$$\{ab(o) \leftrightarrow \mathsf{ctxt}(c(o)) \vee \mathsf{U}\}$$

Since $\mathsf{ctxt}(c(o))$ is *false*, $ab(o)$ remains *unknown* under Ł-logic despite the use of the context operator.

Contextual logic programs are introduced for modelling the syllogistic reasoning task, because otherwise the *contraposition* principle (introduced in Section 3.3.2) could not be applied. Reconsider the syllogism AO3 as a motivating example:

All a are b. (Aab)
Some c are not b. (Ocb)

These premises are represented by the logic program $\mathcal{P}_{AO3}$ defined in Section 3.1. The least model of the weak completion of $\mathcal{P}_{AO3}$, $\langle I^\top, I^\bot \rangle$, is:

$$\langle \{a(o_1), b(o_1), c(o_2), c(o_3), \boxed{b'(o_2)}\},$$
$$\{ab_{ab}(o_1), ab_{ab}(o_2), ab_{ab}(o_3), ab_{cnb}(o_2), ab_{nbb}(o_2), ab_{nbb}(o_3)\}\rangle$$

Note that $b'(o_2) \in I^\top$ (highlighted in gray), but not $b(o_2) \in I^\bot$, even though being able to make conclusions about negated atoms is the whole purpose of the *negation by transformation* principle. Furthermore, since $\exists X : c(X) \in I^\top$, but $\neg\exists X : b(X) \in I^\bot$, the premise Ocb does not even follow from the least model!

It is obvious that the premises of a syllogism must be entailed by the least model of the logic program representing it, especially if any conclusions made by the participants shall be predicted.

The reason for the unexpected least model in this case is the following clause:

$$b(X) \leftarrow a(X) \wedge \neg ab_{ab}(X)$$

Since $a(o_2)$ is *unknown* and $ab_{ab}(o_2)$ is *false*, the whole body of the clause is *unknown* as well. However, in order for $b(o_2)$ to be *false* under $\models_{wcs}$, the body of *all* clauses with $b(o_2)$ in the head must be *false*. Since $b'(o_2) \in I^\top$ already indicates that $b(o_2)$ should be *false*, it is more or less a technical problem that has to be solved.

The *conditional by licenses* principle that generates the clause in question states that

for all X, $b(X)$ holds if $a(X)$ holds and nothing abnormal is known.

Here, it can be considered an abnormality that $b'(X)$ is already *true*, since concluding $b(X)$ afterwards would be a contradiction. Therefore, $ab_{ab}(o_2)$ must be set to *true*, i.e. the negative assumption $ab_{ab}(X) \leftarrow \bot$ must be defeated for o_2.

If the assumption would simply be defeated by adding the clause $ab_{ab}(X) \leftarrow b'(X)$ to the program, $ab_{ab}(X)$ would become *unknown* in cases where nothing is said about $b'(X)$. This would have negative impacts on other syllogism by restricting certain conclusions. The *context* operator allows defeating negative assumptions without affecting such cases, because $\mathsf{ctxt}(L) = \bot$ if $I(L) = \mathsf{U}$. For premises with mood A, it is sufficient to add the following clause to the program:

$$ab_{ab}(X) \leftarrow \mathsf{ctxt}(b'(X))$$

In the weak completion, it is combined with the assumption as follows:

$$ab_{ab}(X) \leftrightarrow \mathsf{ctxt}(b'(X)) \vee \bot$$

which is logically equivalent to:

$$ab_{ab}(X) \leftrightarrow \mathsf{ctxt}(b'(X))$$

Finally, this allows $b(X)$ to be *false* if $b'(X)$ is *true* and leaves all other cases unchanged.

The least model of the weak completion of $\mathcal{P}^2_{AO3} = \mathcal{P}_{AO3} \cup \{ab_{ab}(X) \leftarrow \mathsf{ctxt}(b'(X))\}$ is as follows (the atoms regarding the inference of $\neg b(o_2)$ are highlighted in gray):

$$\langle\{a(o_1), ab_{ab}(o_2), b(o_1), c(o_2), c(o_3), b'(o_2)\},$$
$$\{ab_{ab}(o_1), ab_{ab}(o_3), ab_{cnb}(o_2), ab_{nbb}(o_2), ab_{nbb}(o_3), b(o_2)\}\rangle$$

The premise Ocb follows from the model as intended.

The use of the *context* operator can be generalized to all premises whose mood is A. As expected, the premises then follow from the least model of their logic programs, but no further conclusions are entailed. Instead, the *contraposition* principle becomes applicable.

This approach — defeating the assumptions that the abnormalities of the conditionals are *false* — can be extended to premises of the mood I using the *context* operator:

$$ab_{ab}(X) \leftarrow \mathsf{ctxt}(b'(X)), \qquad \text{(I1)}$$
$$ab_{ab}(o_2) \leftarrow \mathsf{U} \qquad \text{(I2)}$$

where o_2 is the object generated by the *unknown generalization import* principle. The additional clause I2 is necessary for the objects that have been introduced by the principle *unknown generalization*, because otherwise the abnormality may be assumed to be false for this object as well and the premise would effectively be modelled as if it was universally quantified.

Note that using the *context* operator this way has an impact on the conclusions. For some syllogisms that were previously answered with NVC or existentially quantified conclusions (e.g. IIX, $X \in \{1, 2, 3, 4\}$), now conclusions (both existentially or universally quantified) are predicted. It is true that the second added clause (I2) inhibits inferences about the object imported by the premise using the *context* operator. However, inferences about all objects imported by the other premise are still allowed, because for them the assumption is defeated.

This use of the *context* operator is interesting, because many of the new conclusions it leads to are drawn by a significant amount of the participants, while a smaller portion is not. This can be considered as a new principle which shifts the paradigm of the *unknown generalization* principle. Originally, for an existentially quantified principle, inferences about exactly one object were allowed. Now, inferences about exactly one object are explicitly forbidden. Consequently, this more liberal reasoning approach leads to more conclusions. Therefore, we call this principle *deliberate generalization.*

3.3.2 The Contraposition Principle

In classical logic, an implication of the form $a \rightarrow b$ is equivalent to its contrapositive statement $\neg b \rightarrow \neg a$. Therefore, given such an implication, whenever b is known to be *false*, it can be deduced that a is *false* as well. Contraposition also holds under the three-valued Ł-logic, as it is shown in Table 5:

y	z	$y \to z$	$\neg y$	$\neg z$	$\neg z \to \neg y$
$\top$	$\top$	$\top$	$\bot$	$\bot$	$\top$
$\top$	U	U	$\bot$	U	U
$\top$	$\bot$	$\bot$	$\bot$	$\top$	$\bot$
U	$\top$	$\top$	U	$\bot$	$\top$
U	U	$\top$	U	U	$\top$
U	$\bot$	U	U	$\top$	U
$\bot$	$\top$	$\top$	$\top$	$\bot$	$\top$
$\bot$	U	$\top$	$\top$	U	$\top$
$\bot$	$\bot$	$\top$	$\top$	$\top$	$\top$

Table 5: Truth table for contraposition under Ł-logic.

Contraposition can be applied as a principle to the syllogistic reasoning task. Premises with the mood A, e.g. *All a are b*, can logically be seen as an implication of the form $a \to b$. Premises with the mood E, e.g. *No a are b*, correspond to an implication of the form $a \to \neg b$. Their contrapositives are $\neg b \to \neg a$ and $b \to \neg a$, respectively.

Under the Weak Completion Semantics, the premises of a syllogism are modelled as implications with licenses and allow consequences only in one direction. For a premise of the form *All a are b*, no statements about a being *false* can be made, even if b is known to be *false*, because the contraposition is not modelled. However, the data from the meta-analysis by Khemlani and Johnson-Laird [KJL12] gives evidence that some humans use it as a principle of reasoning. Recall the syllogism AO3 as an example: the participants answer both Oca and NVC significantly. Under the Weak Completion Semantics, however, only NVC is predicted. The reason for that can be seen in the least model of the weak completion of the logic program $\mathcal{P}^2_{AO3}$ representing the syllogism with the *deliberate generalization* principle (the relevant atoms are highlighted in gray):

$$\langle\{\, a(o_1)\, , ab_{ab}(o_2), b(o_1),\, c(o_2)\, ,\, c(o_3)\, , b'(o_2)\},$$
$$\{ab_{ab}(o_1), ab_{ab}(o_3), ab_{cnb}(o_2), ab_{nbb}(o_2), ab_{nbb}(o_3), b(o_2)\}\rangle$$

In order to conclude Oca, $a(o_2)$ should be interpreted as *false* (corresponding to $c(o_2)$ from the existential import) and $a(o_3)$ should not be interpreted as *false* (as opposed to $c(o_3)$ from the unknown generalization). That is impossible as long as no rule for $\neg a(X)$ is contained in the logic program. Therefore, the contraposition principle is encoded as follows:

$$a'(X) \leftarrow \neg b(X) \wedge \neg ab_{nba}(X) \qquad \textit{contrapositive rule + licenses}$$
$$a(X) \leftarrow \neg a'(X) \wedge \neg ab_{naa}(X) \qquad \textit{transformation + licenses}$$
$$ab_{nba}(X) \leftarrow \bot \qquad \textit{negative no refutation + licenses}$$

Since $\neg a(X)$ is a negated atom, the *negation by transformation* principle must be used as well. The *no refutation* principle is used because it must be assumed that the abnormality predicate $ab_{nba}(X)$ is *false* for all X for which $b(X)$ is true. It is universally quantified to be consistent with the encoding of the original premise.

We consider a program $\mathcal{P}^3_{AO3}$ which consists of $\mathcal{P}^2_{AO3}$ and the encoding of the *contraposition* principle. The least model of the $\mathcal{P}^3_{AO3}$ contains $a(o_2)$ as a *false* atom, but not $a(o_3)$, as it was intended:

$$\begin{aligned}\langle\{ & a(o_1), ab_{ab}(o_2), b(o_1), c(o_2), c(o_3), a'(o_2), b'(o_2)\},\\ \{ & a(o_2), ab_{ab}(o_1), ab_{ab}(o_3), ab_{cnb}(o_2), ab_{nba}(o_1), ab_{nba}(o_2),\\ & ab_{nba}(o_3), ab_{nbb}(o_2), ab_{nbb}(o_3), b(o_2), a'(o_1)\}\rangle\end{aligned}$$

Oca is the only conclusion entailed by this model. With this approach, two clusters of human reasoners have been identified and can be modelled:

1. people that apply the contraposition principle and entail Oca (ca. 40 %)
2. people that do not apply it, answering NVC (ca. 20 %)

As the answers of the participants of the various studies are accumulated, so are the different predictions of the Weak Completion Semantics. The resulting prediction is that human reasoning processes can lead to the answers Oca and NVC for this particular syllogism, taking into account the individual differences in reasoning.

The contraposition principle is generalized to all syllogisms that have an A mood in one of their premises and a negative mood (E or O) in the other one. As a result, the syllogisms AE3, AO3, and EA3, whose conclusions previously were predicted incompletely by the Weak Completion Semantics, are now solved correctly (perfect match). This does not apply for OA3, however. Obviously, the principle only improves syllogisms of figure 3. This has technical reasons that result from the encoding of premises as implications. The syllogistic figure 3 consists of the premises Xab and Xcb where $X \in \{A, E, I, O\}$. In the logic program, the rules representing both premises have $b(X)$ in the head, so neither $a(X)$ nor $c(X)$ can be entailed for any object, except for the ones imported. Consequently, there can be no conclusion for any of these syllogisms. The contraposition principle introduces a rule with either $a(X)$ or $c(X)$ in the head enabling conclusions for figure 3. It does not affect figure 4, because it consists of the premises Xba and Xbc ($X \in \{A, E, I, O\}$), so its logic program does not contain rules with $b(X)$ in the head. The contraposition principle adds a rule with $\neg b(X)$ in the body, but since this cannot be contained in the least model for any X, no additional conclusions are possible. For the figures 1 and 2 the premises do not match either, this can be shown using a similar argumentation as above. In summary, the premises "do not match" in a sense that the term in the head of the negative premise is not in the body of the affirmative premise, so no conclusions can be drawn. However, this does not hold if the *converse interpretation* principle is applied, because then a rule with the necessary atom in the head is part of the logic program. As a consequence, the Weak Completion Semantics predicts too many answers for the syllogisms AE1 and EA2 if the contraposition principle is used along with the *converse interpretation* principle. A possible solution is limiting the application of the contraposition principle to figure 3.

As mentioned above, the contraposition is also valid for premises with the mood E. However, this case raises the problem that the contraposition of a premise of the form "No a are b" coincides with its *converse interpretation*. The only difference between the two is the existential import of an object, which also leads to slightly different results.

Although not logically valid, the contraposition can be formulated for premises with I moods in the same way as it is applied to A moods. This leads to different conclusions, some of which are in accordance with the results of [KJL12] and some of which are not. Therefore, we do not assume that contraposition is applied by humans in these cases.

3.4 Heuristic Solving Strategies

Until now we have assumed that humans rely on logic when solving syllogistic reasoning tasks. Due to the use of a non-monotonic, three-valued logic, many conclusions can be drawn that are not possible under classical logic.

There are, however, still conclusions the Weak Completion Semantics cannot predict with any of the known principles. For example, consider the syllogisms EEX ($X \in \{1, 2, 3, 4\}$): while nothing can be concluded from only negative premises, a significant amount of participants still answers with Eac or Eca.

When assuming clusters of human reasoners that differ in the principles they apply, it may be reasonable to assume that some of them do not use logic at all. We say that they use certain heuristics, i.e. rules that state which conclusion to choose based on the appearance of the premises. This section introduces two such solving strategies that are well-known in psychology and one that results from own observation.

3.4.1 The Atmosphere Effect

The atmosphere effect was introduced in an experimental study by Woodworth and Sells in 1936 [WS35]. They state that instead of applying logic, humans can heuristically draw conclusions based on the moods of the given premises:

- If both premises have the same mood, the conclusion is likely to have it as well
- If the moods of the premises differ, then
  1. if one mood is negative, the conclusion is very unlikely to be positive (due to the 'negative atmosphere')
  2. if one mood is existentially quantified, the conclusion is very unlikely to be universally quantified (due to the 'existential atmosphere')

The hypothesis was tested in a study with 65 adults that had no training in logic. They were given pairs of premises with a conclusion and should decide whether that conclusion validly followed from the premises. 169 syllogisms were used, but only 42 of them were valid. For the remaining cases, if participants accepted invalid conclusions, it was tested whether they preferred the ones predicted by the atmosphere effect. The authors stated that the experimental data supported the hypothesis. However, Wetherick and Gilhooly have pointed out problems in the implementation of the study. Instead of the originally reported results, it is probable that some participants used logic while others applied an arbitrary heuristic strategy [WG95].

Application to the Weak Completion Semantics There are two possibilities to use the atmosphere effect for modelling clusters of human reasoners.

Generative approach Model that an individual solves a syllogism heuristically and selects a conclusion based on the premises' atmosphere.

Filtering approach Model that conclusions that are unlikely due to the atmosphere hypothesis are not selected.

The *generative approach* is based on the assumption that a person might decide to not apply logic at all if they have to solve a difficult reasoning task. Their evaluation process must be modelled differently from reasoners. Instead of constructing a model, a random conclusion that is likely under the atmosphere hypothesis is selected. By doing so, a new cluster is formed: Guessers using this particular heuristics.

Whether or not an individual is likely to use logic or a heuristic strategy can be expressed by a probabilistic decision. The probability can be trained from experimental data acquired in psychological studies using an algorithm like Expectation-Maximization. As an example on how this can be done, see the parameter training of Multinomial Processing Trees [HB94].

The *filtering approach* aims at improving the combined predictions of the Weak Completion Semantics rather than explaining individual reasoners. It is assumed that a significant subset of the participants of psychological studies indeed uses certain heuristics instead of logic. Therefore, if a conclusion predicted by the Weak Completion Semantics is unlikely under the atmosphere hypothesis, a significant amount of people would not draw it. Since they have to give an answer, they select one according to the heuristics they use or NVC.

The filter is implemented after the conclusions have been drawn from the least model of the weak completion of the logic program representing the syllogisms. It checks if these conclusions conflict with the atmosphere of the premise. If so, NVC is added as the answer that would likely be given by the cluster that does not use logic.

However, tests have shown that the atmosphere effect as a general filter for conclusions is unsuitable to improve the predictions of the Weak Completion Semantics. In the basic principles, it is already implemented that no conclusions conflict with the atmosphere hypothesis. To illustrate this, an example for each case is presented below.

The fact that negative moods create a negative atmosphere (in which affirmative conclusions are not drawn) follows from their representation via the *negation by trans-*

formation principle (e.g. in the premise Eab):

$$
\begin{array}{ll}
b'(X) \leftarrow a(X) \wedge \neg ab_{nab}(X) & \textit{negative rule + licenses} \\
b(X) \leftarrow \neg b'(X) \wedge \neg ab_{nbb}(X) & \textit{transformation + licenses} \\
ab_{nab}(o_1) \leftarrow \bot & \textit{existential import + licenses} \\
a(o_1) \leftarrow \top & \textit{existential import} \\
ab_{nab}(X) \leftarrow \bot & \textit{negative no refutation + licenses}
\end{array}
$$

Here, $a(o_1)$ is *true*, but $b(o_1)$ cannot be *true*. The other, affirmative, premise cannot infer $c(o_1)$ to be *true* either, because the *quantified assertion as conditional* principle relies on $b(o_2)$ being *true* for inference:

$$c(X) \leftarrow b(X) \wedge \neg ab_{bc}(X)$$

Therefore, no affirmative conclusions about a and c are possible. An affirmative conclusion in the opposite direction (i.e., about c and a) is not possible either, because in the original case there is no rule with an a predicate in the head, hence the object imported by the second premise can never be an element of a. If the *converse interpretation* principle is used, the added premise will use the *negation by transformation* principle as well, suppressing affirmative conclusions as described above.

Under the Weak Completion Semantics, existentially quantified moods suppress universally quantified conclusions by the *unknown generalization* principle (e.g. in the premise Iab):

$$
\begin{array}{ll}
b(X) \leftarrow a(X) \wedge \neg ab_{ab}(X) & \textit{rule + licenses} \\
ab_{ab}(o_1) \leftarrow \bot & \textit{existential import + licenses} \\
a(o_1) \leftarrow \top & \textit{existential import} \\
a(o_2) \leftarrow \top & \textit{unknown generalization}
\end{array}
$$

Here, $a(o_2)$ is *true*, but $b(o_2)$ is always *unknown*. The other premise cannot infer $c(o_2)$ to be *true*, because the *conditionals as licences* principle relies on $b(o_2)$ to be *true* for that (compare with above). Therefore, no universal conclusions about a and c are possible. Using the same argument as above, it is obvious that universal conclusion in the opposite direction are not possible either.

The remaining statement of the atmosphere hypothesis, i.e. identical moods in the premises likely lead to a conclusion of the same mood, is easy to show. For the mood O it follows trivially from the arguments made above. For the moods A and E, the conclusions I and O, resp., are excluded by the *no refutation by counterexample* principle: all objects that belong to one category automatically belong to the other, allowing only universally quantified conclusions. For the moods A and I negated atoms do not occur in the logic program, hence the conclusions E and O are impossible. By that, it has been shown for all cases of identical moods in the premises that no conclusion predicted by the Weak Completion Semantics can have a different mood than the premises. Additionally, we may conclude that there is some logic underlying the atmosphere hypothesis, although it is a heuristic strategy. The conclusions that are unlikely due to the atmosphere are not drawn by our logic approach either. However, atmosphere permits many conclusions that do not result from logic.

3.4.2 The Matching Strategy

The matching strategy was introduced by Wetherick and Gilhooly [WG95] as an alternative heuristics to the atmosphere hypothesis. It defines the following order of 'conservatism' on moods:

$$A < I = O < E$$

The higher a mood is in this order, the smaller is the number of entities it makes an assertion about; the mood is 'more conservative'. A conclusion with a certain mood cannot be drawn if one of the premises has a higher mood with respect to that order. People would reason 'conservatively' in the sense that the conclusions they draw do not cover larger sets of objects than the premises.

Wetherick and Gilhooly [WG90] investigated the results of Woodworth and Sells [WS35] to show that there are two groups of reasoners: one that uses logic and one that uses the matching strategy. 23 of the 64 syllogisms have valid conclusions and the matching strategy correctly predicts the conclusions of 14 of these 23 syllogisms. Here it is impossible to say whether a person giving the correct answer used logic or a heuristic strategy. For the remaining 13 syllogisms, if a person gives the correct answer they cannot have used the matching strategy. Wetherick and Gilhooly claim that such individuals would not use the matching strategy for other syllogisms either. They were able to justify this hypothesis by correlating the answers to a subset of valid syllogisms which are an obvious case for matching, but are predicted wrongly by the matching strategy, with the answers to the remaining valid syllogisms in the study by Woodworth and Sells.

Application to the Weak Completion Semantics Similar to the atmosphere effect, we test whether these findings can be used to improve the predictions of the Weak Completion Semantics. The same approaches are used, the only difference is that the predictions of the matching strategy are considered instead of the atmosphere.

It is again suggested to model individual reasoners applying heuristics instead of logic using the *generative* approach. However, it is more interesting that in contrast to the atmosphere hypothesis, the *filtering* approach can have an effect on the conclusions depending on the principles used.

If the *converse interpretation* principle is not applied or only applied for the I premise, the answers does not change at all. If it is applied to the E premise as well, there are cases where conclusions are predicted which conflict with the matching strategy, namely the IEX and EIX syllogisms ($X \in \{1, 2, 3, 4\}$). These syllogisms are special in the sense that in many cases the valid conclusion Oac (or Oca, resp.) and in most cases NVC has been chosen by a significant amount of the participants, and these answers often overlap. When assuming two clusters of reasoners, the experimental data is matched best. Note that Eac and Eca are also significant answers for these syllogisms. They can be explained as guesses according to the matching strategy (atmosphere fails in this case, because the premise with mood I creates an existential atmosphere). In contrast to the atmosphere hypothesis, the matching strategy leads to a significantly better fit of the predictions of the Weak Completion Semantics to the answers of the participants presented in [KJL12].

3.4.3 Biased Conclusions for Syllogisms of Figure 1

An in-depth analysis of the participants' answers to syllogisms with figure 1 leads to the following observation:

- for a syllogistic premise $XY1$ ($X, Y \in \{A, I, E, O\}$), an answer Zac is always given by a significant amount of participants
  - Z is the highest possible mood with respect to the order defined by the matching strategy, given the moods X and Y
  - if both the mood I and O are possible, O is preferred
- only if Zac is an invalid conclusion, other (but not necessarily correct) answers are given

For illustration, consider the syllogism AE1:

All a are b.
No b are c.

It has Eac and Eca as valid conclusions (Oac and Oca as subsets are neglected). The participants only concluded Eac. Compare this with the syllogism EA1:

No a are b.
All b are c.

It has only Oca as a valid conclusion, but is answered with Eac, too, and nothing else. An example for an additional answer is the syllogism EI1:

No a are b.
Some b are c.

The majority of the participants answered Eac, which is incorrect. NVC was also a significant answer, though Oca is the only valid conclusion.

The effect described above is true for 15 out of the 16 syllogisms with figure 1 (OE1 being the exception), so it is considered to be significant. A possible explanation for this observation may be given by the difficulty of the reasoning process. It has been shown [Dic78] that the figure of a syllogism determines its difficulty to a larger degree. Syllogisms with figure 1 only involve forward reasoning and are thus very easy to solve. We assume that in such situations many humans do not even start reason about the syllogism and just give a rash answer, because the task seems to be so simple.

These findings can again be taken into account when modelling the syllogistic reasoning task in two ways. We use an approach similar to the *filtering approach* for the heuristics introduced earlier. Since so many people seem to solve syllogisms with figure 1 heuristically, all answers conflicting with the *biased conclusions* are suppressed (except NVC). Instead, the heuristic conclusion is answered. In a certain sense, this approach is also *generative*. However, since there is only one answer for each syllogism under this particular heuristics, it can be modelled without assuming a probabilistic process.

4 Representation of Reasoning Processes

Until now, the actual reasoning process of a human has not played an important role. In the last section, it has been shown how individual reasoners can been modelled by turning on or off certain principles. The conclusions drawn from each of the corresponding logic programs have been combined to a single prediction. While this is suitable for showing how good the theory fits the participants' answers, the information on what reasoning process leads to what conclusions gets lost.

This chapter presents an approach that is able to show how the individual reasoning processes differ and the effects these differences have on the conclusions drawn. This is achieved by using tree models that contain each possible reasoning process as a path from the root to a leaf containing the conclusion.

Binary decision trees and multinomial processing trees have been selected as frameworks for modelling the syllogistic reasoning task for the following reasons: decision trees are well-known and widely spread in the computer science community, so much support for creating the models and training them is available. Multinomial processing trees have recently been suggested [RSS14] as cognitive models in order to make different cognitive theories comparable.

4.1 Decision Trees

Definition 23 (Decision Tree) *A decision tree is a tree whose nodes have a special semantics:*

Internal node *decision at a certain point in the decision process*

Leaf *outcome of the decision process if exactly the decisions in the nodes on the branch from the root to the leaf are made*

A decision tree is binary *if each inner node has exactly two successors, i.e. each decision is between two alternatives. Each decision tree can be transformed into a binary decision tree.*

For modelling reasoning processes, the outcome of the decision process is what conclusions should be given as an answer. Each decision is the question whether or not a certain principle should be applied. Note that these decisions are always binary. Each branch of the decision tree corresponds to a possible reasoning process of an individual, using exactly the principles for which a positive decision was made.

Note that the branches of such a tree create the impression of an implicit order of making decisions. For two consecutive nodes in the tree, it seems that the decision represented by the first node is also made first by the reasoner. This is wrong, because it is not the goal to model the reasoning process like an algorithm, only what steps actually take place. Therefore, the chronological order of the reasoning process is unknown and the decision nodes might be changed in order. This becomes clear when considering the semantics of the tree: each branch is encoded as a logic program consisting of a common part (principles used by all reasoners) and the encoding of all principles for which a

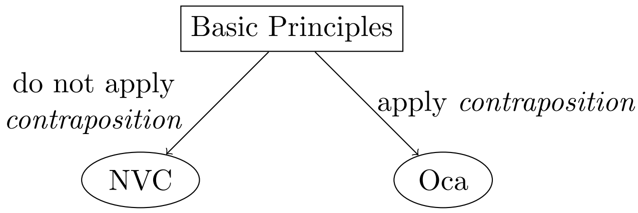

Figure 1: Decision tree for the syllogism AO3

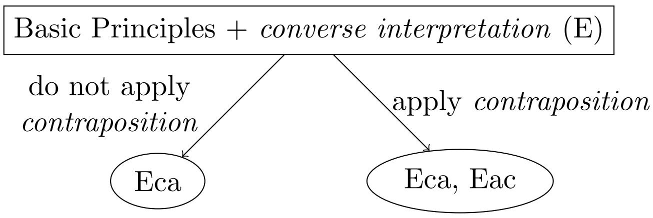

Figure 2: Decision tree for the syllogism EA3 (converse interpretation for E assumed)

positive decision was made. The conclusions entailed by the least model of this logic program form the leaf of the branch. Since the least model of the weak completion of a logic program does not depend on the order of the rules, the predicted conclusions are invariant with respect to the order of decisions.

The common part of the logic program (*basic program*) consists of an encoding of both premises with the basic principles that are necessary for the particular mood. For a single syllogism, other principles may be added via decision nodes. If adding a principle would lead to conclusions that are not drawn by the participants, it may be omitted. If a decision against a principle would lead to conclusions that are not drawn by the participants, its encoding may be added to the basic logic program without introducing a decision node. As an example, see the decision tree for the syllogism AO3 in Figure 1. It consists of the basic principles for all syllogisms and the option to apply the contraposition principle, all leaves explain the significant answers by the participants.

In many cases, the Weak Completion Semantics predicts more than one conclusion for the principles described by a branch in the decision tree. While leaves can contain sets of conclusions from a theoretical point of view, in many experiments, the participants are only allowed to give one answer. In order to model this, additional decision nodes representing the choice between possible alternatives are introduced. See Figure 2 and Figure 3 for comparison.

As it is possible that the prediction without a principle is a subset of the prediction with that principle applied, the decision tree may become a directed acyclic graph (DAG), cf. Figure 3. There are two possibilities to handle this:

1. Relabel the leaves with the same conclusions, e.g. to 'Eca without contraposition' and 'Eca by choice' for the syllogism EA3 in Figure 3.

2. Instead of branches from the root to a leaf, consider paths from the source to a

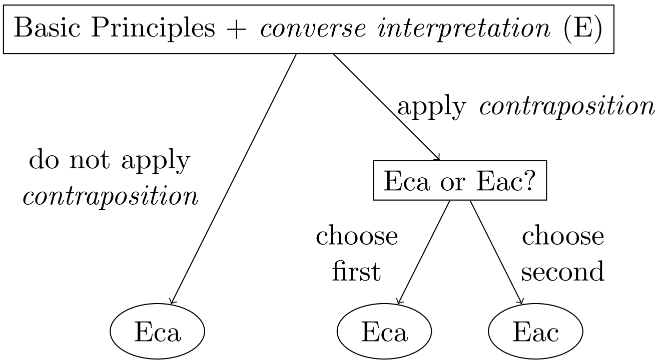

Figure 3: Decision tree for the syllogism EA3 (converse interpretation for E assumed) with choice between alternatives

sink. Since there is exactly one source (the former root) and the only nodes with multiple ingoing edges are the former leaves, the DAG is still suitable for modelling decisions.

The goal for a decision tree is that it contains a branch for each answer given by a participant, the principles used in this branch then explain the reasoning process of the person. Since at the time of writing only accumulated data from studies are available, the goal is that the tree's branches explain exactly the significant answers given by the participants.

If for each of the 64 syllogisms a decision tree is build by hand, the result may be a perfect match of the participants' answers that is subject to overfitting to a large degree. Instead, the goal should be a theory for generating the trees uniformly based on figure and mood of the premises.

4.2 Multinomial Processing Trees

Multinomial processing trees are an established model for modelling cognitive processes [RB88] that is also very well suited for quantitative methods from statistics and computer science. Recently, they have been introduced as a generalizing approach for cognitive theories to model the syllogistic reasoning task [RSS14].

Definition 24 (Multinomial Processing Tree) *A Multinomial Processing Tree (MPT) is a directed acyclic graph with a finite set of response categories as leaves (actually sinks) and a finite set of cognitive processes as inner nodes. Each edge has a parameter assigned corresponding to the probability that its preceding process node is followed by its successor.*

A MPT for a syllogism is similar to the corresponding decision tree. Each inner node represents a principle that may or may not be used by an individual reasoner. Only binary MPTs are considered, so that the 'outcome' of the process represented by the node is just whether the principle was used or not.

The probabilistic parameters are assigned to each outgoing edge of a node as follows: if the process represented by the node leads to using a principle with probability p, then the positive outgoing edge is assigned the parameter p and the negative outgoing edge is assigned the parameter $1 - p$. These probabilities can be learnt from experimental data using algorithms such as Expectation-Maximization [HB94].

Following a branch of a MPT from the source to the sink the reasoning process represented by this branch is obtained. As for decision trees, the actual order of nodes does not have a meaning. The answer in the leave is just a conclusion entailed by the least model of the weak completion logic program that encodes all principles used at this branch. The probability of a branch is the product of the parameter values at its edges. The probability of an answer is the sum of the probabilities of all branches leading to it.

The goal of building a MPT is that the distribution of answers predicted by it deviates from the prediction of answers among the participants as little as possible. Various quantitative quality measures exist for MPTs, for details consider [Aka74, $\mathrm{S}^{+}78$].

For use in quantitative models, a MPT must have exactly 9 sinks, one for each of the possible conclusions and NVC[3]. Leaves must not contain multiple conclusions; such predictions must be split by introducing choice decisions as shown for decision trees.

Guessing Trees The Weak Completion Semantics cannot explain all 9 answers for each syllogism. Consequently, many MPTs would have missing sinks and would not meet the requirement from above. In many cases, these missing conclusions cannot by entailed by logic and are not given by a significant amount of participants. In some cases, however, the reason is that the participants apply a principle yet unknown or do not use logic at all. As a solution for both situations, guessing trees are introduced.

Definition 25 (Guessing Tree) *A guessing tree is a MPT whose nodes are not cognitive processes, but rather stochastic trials determining the path to be taken. The leaves are the set of conclusions out of which a guess is made.*

Guessing can be *totally random*, allowing all conclusions, or *educated*, allowing only the conclusions predicted by a heuristic strategy, e.g. matching. See Figure 4 as an example for a guessing tree under the matching strategy whose corresponding syllogism does not have a premise with the mood E. Among others, this guessing tree is appropriate for the syllogism AO3.

Since some individuals may use logic while others guess in the same syllogistic reasoning task, the guessing tree is combined with the reasoning tree by adding a new source node with both former trees as successors. The meaning of the source node is to represent the distribution between reasoners and guessers among the participants. This can again be modelled as a stochastic trial. See Figure 5 as a possible MPT for the syllogism OI1, where Iac is a significant answer that is not predicted by the Weak Completion Semantics. This shows how the syllogism can be modelled using a both logic and heuristic

[3]Although sets of conclusions are theoretically possible, they are impractical: there are 2^9 possible answers, most of which are not present in the training data. The quantitative methods used to train MPTs do not produce good results if the contribution is that sparse.

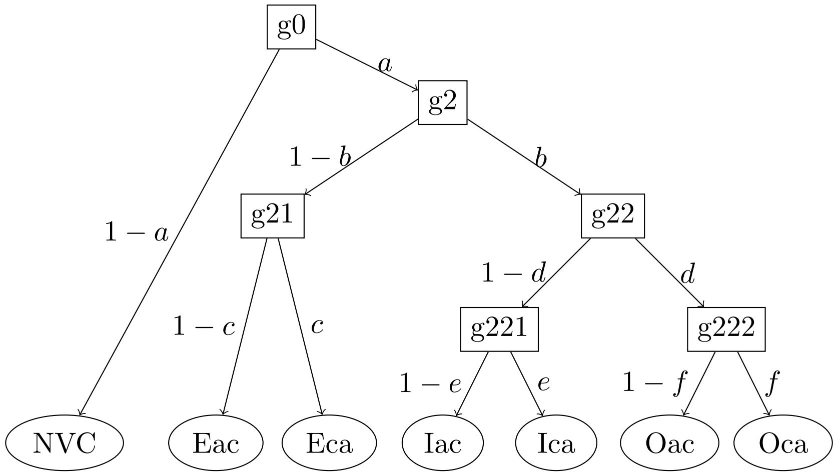

Figure 4: Example MPT for guessing under the matching strategy

strategies, although we will later only use the *biased conclusions in figure 1* heuristics to model it.

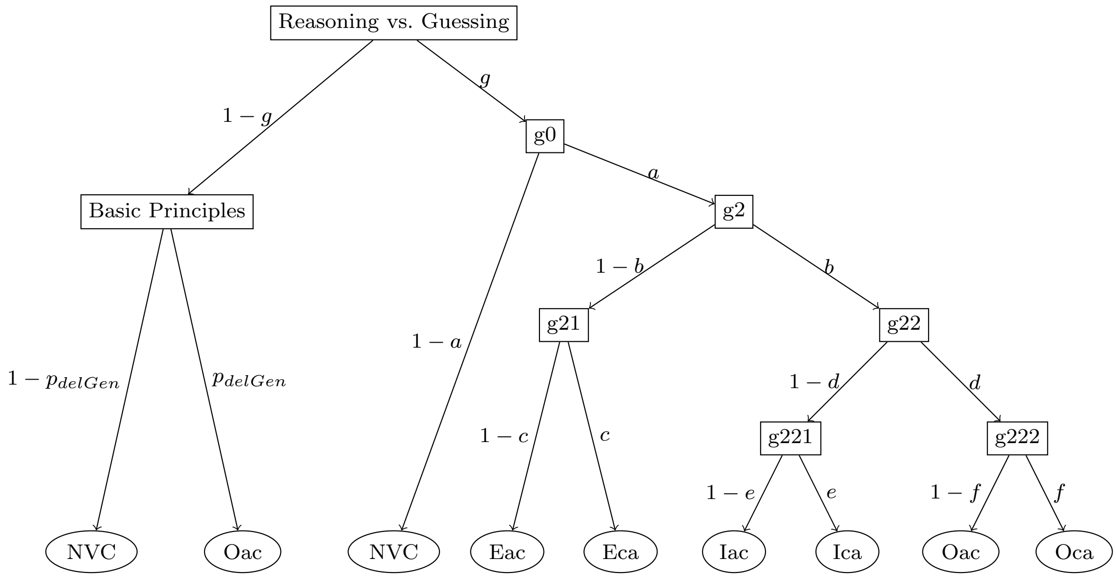

Figure 5: MPT for the syllogism OI1 (p_{delGen} corresponds to the probability of applying the *deliberate generalization* principle).

5 Evaluation

In this section it will be shown how the accuracy of predictions of cognitive theories with respect to experimental data is calculated. After that, it will be presented how well the Weak Completion Semantics fits the data accumulated in the meta-analysis on syllogisms [KJL12] and how it performs compared to other cognitive theories.

5.1 Accuracy of Predictions

Khemlani and Johnson-Laird [KJL12] used the following representation of answers to syllogistic reasoning tasks: for each of the 64 syllogisms, give the answer as a vector $answer \in \{0, 1\}^9$. The nine positions of the vector correspond to the possible conclusions in the order Aac, Eac, Iac, Oac, Aca, Eca, Ica, Oca, and NVC. Each position of the vector contains a 1 if and only if the corresponding answer is given, and a 0 otherwise.

Such vectors are created for both the answers of the participants of the studies and the predictions of the cognitive theory. However, the answers of the participants are accumulated data; for each possible conclusion the percentage of participants answering it is given. Therefore, Khemlani and Johnson-Laird have introduced a threshold of statistical significance. Conclusions that have been given by more than 16 % of the participants are assigned a 1 and others are assigned a 0. The accuracy of the predictions is based on the Hamming distance between the two vectors. Intuitively, a single answer is rewarded if the participants and the theory coincide. The predicted answers for a

Principle of reasoning	Applicability
Quantified assertion as conditional	All syllogisms
Licenses for inferences	All syllogisms
Existential import	All syllogisms
Unknown generalization	Premises with an existential mood
No refutation	Premises with a universal mood
Negation by transformation	Premises with a negative mood
No derivation by double negation	Premises with a negative mood
Converse interpretation	Premises with the mood E or I
Deliberate generalization	Premises with mood I
Contraposition	Premises with a universal mood
Matching strategy	All syllogisms
Biased conclusions in figure 1	Premises with figure 1

Table 6: Reasoning principles under the Weak Completion Semantics.

syllogism are scored as follows:

$$\mathsf{score}(v_{theory}, v_{participants}) = \frac{1}{9} \times \sum_{i=1}^{9} [v_{theory}(i) = v_{participants}(i)]$$

The division by 9 is performed to obtain the percentage of matching numbers.

The accuracy of predictions for all 64 syllogisms is the average over the scores of each single prediction.

5.2 Results of the Weak Completion Semantics

Table 6 gives an overview on all principles that have been introduced in this and previous work [dCSH17]. The principles *quantified assertion as conditional*, *licenses for inferences*, *existential import*, *unknown generalization*, *no refutation*, *negation by transformation*, *no derivation by double negation*, and the *converse interpretation* of premises with mood I are regarded as *basic principles* and assumed to be used by every reasoner where applicable. The remaining principles are the foundation for modelling clusters.

When clusters of reasoners are to be introduced, the first decision is the number of clusters to model. The second decision is for each cluster, which subset of principles should define that cluster. The quality of such a clustering is then evaluated as described above. However, the combinatorial explosion when considering all possible subsets of principles makes this optimization problem difficult to solve. Therefore, only the best clustering approach that has been found so far is presented here.

The optimal clustering consists of three clusters of reasoners using logic and two clusters of heuristic strategies. The clusters are defined as follows:

Basic: basic principles

Deliberate Generalitaion: basic principles + *deliberate generalization*

Contraposition + Converse E: basic principles + *converse rule* in premises with mood E + *contraposition* in premises with mood A

Matching filter the answers of the Weak Completion Semantics according to the matching strategy

Biases conclusions in figure 1 filter the answers of the Weak Completion Semantics according to the biased conclusions

Note that there are some differences from the definitions from above. The *deliberate generalization* principle is used for premises with mood A in the cluster using *contraposition*. This does not allow new conclusion by itself, but is necessary to be able to draw conclusions from the contrapositive. The *contraposition* is only used for premises with mood A. Premises with mood E are handled by the *converse rule*, which gives very similar, but slightly better results.

Abduction is not modelled as a principle, because it is a different form of reasoning and the assumptions made for deduction may hold for it. It is assumed to be done by all reasoners uniformly.

The heuristic strategies that are modelled are the matching strategy and the biased conclusions in syllogisms with figure 1. Matching is modelled by applying the *filtering approach* to the answers of each cluster; each answer not predicted by matching is suppressed and NVC is answered if no other conclusions remain. The results would be slightly better (ca. 0.3 %) if for the *contraposition + converse E* cluster the original answers were kept. However, this has not been done because of the danger to become subject to overfitting. The bias in figure 1 is implemented according to the approach described in Section 3.4.3.

This clustering is backed by achieving an accuracy of 92.2 %. 32 of the syllogisms are solved perfectly, 20 have one wrong prediction, 11 have two wrong predictions, and OE1 is the worst case with three mismatches. A detailed comparison of the predictions with the answers of the humans is presented in Table 7. The accuracy of the Weak Completion Semantics is significantly higher than those of other cognitive theories, such as the Mental Models theory (78 %) and the Verbal Models theory (84 %).

Syllogism	Premises	Aac	Eac	Iac	Oac	Aca	Eca	Ica	Oca	NVC	Match
AA1	Aab, Abc	1	0	0	0	0	0	0	0	0	100 %
AA2	Aba, Acb	0	0	0	0	1	0	0	0	0	89 %
AA3	Aab, Acb	0	0	0	0	0	0	0	0	1	89 %
AA4	Aba, Acb	1	0	0	0	1	0	0	0	0	78 %
AI1	Aab, Ibc	0	0	1	0	0	0	0	0	0	100 %
AI2	Aba, Icb	0	0	1	0	0	0	1	0	0	100 %
AI3	Aab, Icb	0	0	1	0	0	0	1	0	0	78 %
AI4	Aba, Icb	0	0	1	0	0	0	1	0	0	100 %
AE1	Aab, Ebc	0	1	0	0	0	0	0	0	0	100 %
AE2	Aba, Ecb	0	1	0	0	0	1	0	0	0	100 %
AE3	Aab, Ecb	0	1	0	0	0	1	0	0	1	89 %
AE4	Aba, Ecb	0	1	0	0	0	1	0	0	0	78 %
AO1	Aab, Obc	0	0	0	1	0	0	0	0	0	100 %
AO2	Aba, Ocb	0	0	0	0	0	0	0	1	0	78 %
AO3	Aab, Ocb	0	0	0	0	0	0	0	1	1	100 %
AO4	Aba, Ocb	0	0	0	1	0	0	0	0	0	100 %
IA1	Iab, Abc	0	0	1	0	0	0	0	0	0	100 %
IA2	Iba, Acb	0	0	1	0	0	0	1	0	0	100 %
IA3	Iab, Acb	0	0	1	0	0	0	1	0	0	78 %
IA4	Iba, Acb	0	0	1	0	0	0	1	0	0	100 %
II1	Iab, Ibc	0	0	1	0	0	0	0	0	1	100 %
II2	Iba, Icb	0	0	1	0	0	0	1	0	1	100 %
II3	Iab, Icb	0	0	1	0	0	0	1	0	1	89 %
II4	Iba, Icb	0	0	1	0	0	0	1	0	1	89 %
IE1	Iab, Ebc	0	1	0	0	0	0	0	0	1	78 %
IE2	Iba, Ecb	0	0	0	0	0	1	0	0	1	100 %
IE3	Iab, Ecb	0	0	0	0	0	1	0	0	1	89 %
IE4	Iba, Ecb	0	0	0	1	0	0	0	0	1	100 %
IO1	Iab, Obc	0	0	0	1	0	0	0	0	1	100 %
IO2	Iba, Ocb	0	0	0	0	0	0	0	1	1	89 %
IO3	Iab, Ocb	0	0	0	0	0	0	0	1	1	89 %
IO4	Iba, Ocb	0	0	0	1	0	0	0	0	1	100 %
EA1	Eab, Abc	0	1	0	0	0	0	0	0	0	100 %
EA2	Eba, Acb	0	0	0	0	0	1	0	0	0	89 %
EA3	Eab, Acb	0	0	0	0	0	1	0	0	1	78 %
EA4	Eba, Acb	0	1	0	0	0	1	0	0	0	100 %
EI1	Eab, Ibc	0	1	0	0	0	0	0	0	1	100 %
EI2	Eba, Icb	0	0	0	0	0	0	0	1	1	100 %
EI3	Eab, Icb	0	1	0	0	0	0	0	0	1	89 %
EI4	Eba, Icb	0	0	0	0	0	0	0	1	1	78 %
EE1	Eab, Ebc	0	1	0	0	0	0	0	0	1	100 %
EE2	Eba, Ecb	0	0	0	0	0	0	0	0	1	78 %
EE3	Eab, Ecb	0	0	0	0	0	0	0	0	1	89 %
EE4	Eba, Ecb	0	0	0	0	0	0	0	0	1	89 %
EO1	Eab, Obc	0	1	0	0	0	0	0	0	1	89 %
EO2	Eba, Ocb	0	0	0	0	0	0	0	0	1	100 %
EO3	Eab, Ocb	0	0	0	0	0	0	0	0	1	89 %
EO4	Eba, Ocb	0	0	0	0	0	0	0	0	1	100 %
OA1	Oab, Abc	0	0	0	1	0	0	0	0	0	78 %
OA2	Oba, Acb	0	0	0	0	0	0	0	1	0	100 %
OA3	Oab, Acb	0	0	0	0	0	0	0	0	1	78 %
OA4	Oba, Acb	0	0	0	0	0	0	0	1	0	100 %
OI1	Oab, Ibc	0	0	0	1	0	0	0	0	1	89 %
OI2	Oba, Icb	0	0	0	0	0	0	0	1	1	100 %
OI3	Oab, Icb	0	0	0	1	0	0	0	0	1	100 %
OI4	Oba, Icb	0	0	0	0	0	0	0	1	1	89 %
OE1	Oab, Ebc	0	1	0	0	0	0	0	0	1	67 %
OE2	Oba, Ecb	0	0	0	0	0	0	0	0	1	89 %
OE3	Oab, Ecb	0	0	0	0	0	0	0	0	1	100 %
OE4	Oba, Ecb	0	0	0	0	0	0	0	0	1	100 %
OO1	Oab, Obc	0	0	0	1	0	0	0	0	1	100 %
OO2	Oba, Ocb	0	0	0	0	0	0	0	0	1	89 %
OO3	Oab, Ocb	0	0	0	0	0	0	0	0	1	89 %
OO4	Oba, Ocb	0	0	0	0	0	0	0	0	1	89 %
Overall											92 %

Table 7: Predictions of the Weak Completion Semantics for each syllogism in the clustering approach. Matches with the participants' data are highlighted in light gray, mismatches in dark gray.

6 Conclusion

The cognitive theory based on the Weak Completion Semantics has already been applied to the suppression task [DHR12], the selection task [DHR13], the belief-bias effect [PDH14a, PDH14b], reasoning about conditionals [DH15, DHP15], spatial reasoning [DHH15], and syllogistic reasoning [Die15, CDHR16]. Recently, a general monadic reasoning theory has been proposed and applied to the syllogistic reasoning task [dCSH17]. The theory is modular in the sense that several principles of reasoning have been identified. They allow the encoding of monadic quantified assertions as logic programs and reasoning on them under the Weak Completion Semantics.

Until now, the theory did not consider individual differences in reasoning, nor did any of the cognitive theories compared in the meta-analysis of syllogistic reasoning by Khemlani and Johnson-Laird [KJL12]. However, when looking at the experimental data provided by the study, such differences can be observed. They were accounted in this work by introducing the concept of clusters of reasoners to the Weak Completion Semantics.

Clusters are modelled using the flexible nature of principles. The differences in reasoning are explained by the fact that the application of principles can vary between clusters. This leaves the approach open for the introduction of new principles as the result of future research. Additionally, it has been shown that some individuals might not use logic at all to solve reasoning tasks. Therefore, heuristics that are well-known in psychology, such as the atmosphere hypothesis [WS35] and the matching strategy [WG95] have been analyzed and it has been suggested how to apply them under the Weak Completion Semantics. The clustering approach based on the Weak Completion Semantics achieved an accuracy of 92 % with respect to the answers of the participants reported in [KJL12].

Although the clusters explain the overall answers of humans quite well, they are unsuitable for illustrating what principle enables or inhibits what conclusions in detail. Therefore, tree models have been introduced. Decision trees model to what cluster a human belongs as the decision which principles they apply or do not apply. Multinomial Processing Trees are a probabilistic model that allows quantitative predictions about the distribution of answers.

There are, however, some issues with accumulating the data like Khemlani and Johnson-Laird did. First, the information on individual reasoners gets lost. It is impossible to reconstruct what principles define a cluster from the accumulated data. Second, the accumulation does not make a difference between answers given by 20 % and answers given by 90 % of the participants. It is, however, of particular interest for modelling clusters if a conclusion is drawn by almost all humans or a small, but significant minority. Third, the threshold of 16 % has a high impact on the conclusions drawn by 'the participants'. There are many answers in the study that have been given by an amount of participants close to 16 %. Since the number of participants per study in the meta-analysis is quite small, the significance may vary between studies.

Therefore, it is suggested for future research to concentrate on modelling individual reasoners using e.g. Multinomial Processing Trees. This eliminates the dependence on the 16 % threshold and allows quantified hypothesis about clusters.

References

[Aka74] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974.

[BHN03] Alison Bacon, Simon Handley, and Stephen Newstead. Individual differences in strategies for syllogistic reasoning. *Thinking & reasoning*, 9(2):133–168, 2003.

[CDHR16] Ana Costa, Emmanuelle-Anna Dietz, Steffen Hölldobler, and Marco Ragni. Syllogistic reasoning under the weak completion semantics. 2016.

[Cla78] Keith L Clark. Negation as failure. In *Logic and data bases*, pages 293–322. Springer, 1978.

[dCSH17] Ana Oliveira da Costa, Emmanuelle-Anna Dietz Saldanha, and Steffen Hölldobler. Monadic reasoning using weak completion semantics. 2017.

[DH15] Emmanuelle-Anna Dietz and Steffen Hölldobler. A new computational logic approach to reason with conditionals. In *International Conference on Logic Programming and Nonmonotonic Reasoning*, pages 265–278. Springer, 2015.

[DHH15] Emmanuelle-Anna Dietz, Steffen Hölldobler, and Raphael Höps. A computational logic approach to human spatial reasoning. In *Computational Intelligence, 2015 IEEE Symposium Series on*, pages 1627–1634. IEEE, 2015.

[DHP15] Emmanuelle-Anna Dietz, Steffen Hölldobler, and Luís Moniz Pereira. On conditionals. In *GCAI*, pages 79–92, 2015.

[DHP17] Emmanuelle-Anna Dietz Saldanha, Steffen Hölldobler, and Luís Moniz Pereira. Contextual reasoning: Usually birds can abductively fly. 2017.

[DHR12] Emmanuelle-Anna Dietz, Steffen Hölldobler, and Marco Ragni. A computational logic approach to the suppression task. In *CogSci*, 2012.

[DHR13] Emmanuelle-Anna Dietz, Steffen Hölldobler, and Marco Ragni. A computational logic approach to the abstract and the social case of the selection task. In *Proceedings of the 11th International Symposium on Logical Formalizations of Commonsense Reasoning, COMMONSENSE*, 2013.

[Dic78] Louis S Dickstein. The effect of figure on syllogistic reasoning. *Memory & Cognition*, 6(1):76–83, 1978.

[Die15] Emanuelle-Anna Dietz. A computational logic approach to syllogisms in human reasoning. In *Bridging@ CADE*, pages 17–31, 2015.

[Gri75] H Paul Grice. Logic and conversation. *1975*, pages 41–58, 1975.

[HB94] Xiangen Hu and William H Batchelder. The statistical analysis of general processing tree models with the em algorithm. *Psychometrika*, 59(1):21–47, 1994.

[Höl09] Steffen Hölldobler. *Logik und Logikprogrammierung*. Synchron, Wiss.-Verlag der Autoren, 2009.

[HR09] Steffen Hölldobler and Carroline Dewi Puspa Kencana Ramli. Logic programs under three-valued łukasiewicz semantics. In *International Conference on Logic Programming*, pages 464–478. Springer, 2009.

[JL80] Philip N Johnson-Laird. Mental models in cognitive science. *Cognitive science*, 4(1):71–115, 1980.

[JLK13] PN Johnson-Laird and Sangeet S Khemlani. Toward a unified theory of reasoning. *The psychology of learning and motivation*, 59:1–42, 2013.

[JLS78] Philip N Johnson-Laird and Mark Steedman. The psychology of syllogisms. *Cognitive psychology*, 10(1):64–99, 1978.

[KJL12] Sangeet Khemlani and P. N. Johnson-Laird. Psychological bulletin theories of the syllogism: A meta-analysis. 2012.

[KJL16] Sangeet Khemlani and PN Johnson-Laird. How people differ in syllogistic reasoning. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, 2016.

[Llo87] John Wylie Lloyd. Foundations of logic programming. symbolic computation: Artificial intelligence, 1987.

[Łuk20] Jan Łukasiewicz. O logice trójwartościowej. 1920.

[Łuk57] Jan Łukasiewicz. Aristotle's syllogistic from the standpoint of modern formal logic. 1957.

[PDH14a] Luís Moniz Pereira, Emmanuelle-Anna Dietz, and Steffen Hölldobler. An abductive reasoning approach to the belief bias effect. In *KR*, 2014.

[PDH14b] Luis Moniz Pereira, Emmanuelle-Anna Dietz, and Steffen Hölldobler. Contextual abductive reasoning with side-effects. *Theory and Practice of Logic Programming*, 14(4-5):633–648, 2014.

[Pei74] Charles Sanders Peirce. *Collected papers of charles sanders peirce*, volume 2. Harvard University Press, 1974.

[RB88] David M Riefer and William H Batchelder. Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95(3):318–339, 1988.

[Rip94] Lance J Rips. *The psychology of proof: Deductive reasoning in human thinking*. Mit Press, 1994.

[RSS14] Marco Ragni, Henrik Singmann, and Eva-Maria Steinlein. Theory comparison for generalized quantifiers. In *CogSci*, 2014.

[S+78] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[SMP94] Anthony J Sanford, Linda M Moxey, and Kevin Paterson. Psychological studies of quantifiers. *Journal of Semantics*, 11(3):153–170, 1994.

[SvL08] Keith Stenning and Michiel van Lambalgen. Human reasoning and cognitive science. a bradford book. In *Cambridge MA*. MIT Press, 2008.

[WG90] NE Wetherick and KJ Gilhooly. Syllogistic reasoning: Effects of premise order. *Lines of thinking: Reflections on the psychology of thought*, 1:99–108, 1990.

[WG95] N. E. Wetherick and K. J. Gilhooly. 'atmosphere', matching, and logic in syllogistic reasoning. *Current Psychology*, 14(3):169–178, 1995.

[WS35] R. S. Woodworth and S. B. Sells. An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, 18(4):451, 1935.