Simulating Human Inferences in the Light of New Information: A Formal Analysis

Marco Ragni,[1,†] **Christian Eichhorn,**[2,⋆] and **Gabriele Kern-Isberner**[2,⋆]
⋆ :Technische Universität Dortmund, Dortmund, Germany
† :Technical Faculty, University of Freiburg
[1]ragni@informatik.uni-freiburg.de
[2]{christian.eichhorn,gabriele.kern-isberner}@cs.tu-dortmund.de

Abstract

Human answer patterns in psychological reasoning experiments systematically deviate from predictions of classical logic. When interactions between any artificial reasoning system and humans are necessary this difference can be useful in some cases and lead to problems in other cases. Hence, other approaches than classical logic might be better suited to capture human inference processes. Evaluations are rare of how good such other approaches, e.g., non-monotonic logics, can explain psychological findings. In this article we consider the so-called Suppression Task, a core example in cognitive science about human reasoning that demonstrates that some additional information can lead to the suppression of simple inferences like the modus ponens. The psychological findings for this task have often been replicated and demonstrate a key-effect of human inferences. We analyze inferences of selected formal approaches and compare them by their capacity to cover human inference observed in the Suppression Task. A discussion on formal properties of successful theories conclude the paper.

Introduction

In everyday life we often use conditionals: We use them to explain facts, e.g., "if it rains then the street gets wet", to state predictions about future events, e.g., "if the air pollution continues, the ozone hole increases" or to reason about counterfactuals: "if Oswald had not shot Kennedy, then someone else would have" e.g., [Byrne, 2007]. From a formal logic perspective we focus on *correct inferences* and from a cognitive science/psychology perspective we focus on analyzing *how logically naïve human reasoners* actually do reason. The goal is to develop cognitive models that can predict systematic human reasoning errors. There is a plethora of examples demonstrating that human reasoning does not follow the rules of propositional or first order logic. An example that humans deviate from propositional logic is the famous Wason Selection Task [Wason, 1968]. It demonstrates, in its abstract version, that humans have difficulties with the modus tollens. However, it would be wrong to simply assume that humans make arbitrary errors and may only lack of concentration, motivation, or necessary knowledge. Actually, human, "commonsense", reasoning is very successful in solving most of the problems everybody has to face in ones daily life. Human deviations from classical logic are far from being arbitrary, experiments show systematic deviations of many humans that make similar to identical inferences on the same problems. To explain these findings some researchers turned to probabilistic approaches [Oaksford and Chater, 2007], others to heuristic approaches (for an overview, see [Manktelow, 1999]), and recently others have proposed to analyze other than classical logics [Stenning and Lambalgen, 2008]. The reason for this latter turn is that there are logics that allow for one aspect that propositional logic does certainly not: non-monotonicity. A logic is called *non-monotonic* if additional information can lead to retract previously drawn inferences. In psychology this ability in the human reasoning process is demonstrated with *a core research paradigm*: the Suppression Task [Stenning and Lambalgen, 2008]. Consider the following example taken from [Byrne, 1989] where three groups of participants received one of three types of problems: $\alpha\delta$ (Group 1), $\alpha\beta\delta$ (Group 2), and $\alpha\gamma\delta$ (Group 3):

(β-case) (α) *If she has an essay to write, then she will study late in the library* and — $(e \rightarrow l)$
(β) *If she has a textbook to read, then she will study late in the library* and — $(t \rightarrow l)$
(δ) *She has an essay to write.* — (e)

Most participants (98 %) in the study concluded: *She will study late in the library.* We will write in a succinct formulation for this problem e

(γ-case) (α) *If she has an essay to write, then she will study late in the library* and — $(e \rightarrow l)$
(γ) *If the library stays open, she will study late in the library* and — $(o \rightarrow l)$
(δ) *She has an essay to write.* — (e)

only 38 % of the participants make a modus ponens inference [Byrne, 1989] and conclude in the γ-case that: *She will study late in the library* and 62% concluded that: *She may or may not study late in the library.* This example shows that although the conclusion "she will study late in the library" is still correct, it is suppressed by the γ-conditional, but not by the β- conditional. The effect of the additional β compared to the most simple modus ponens (using $\alpha\delta$) is non existent – both groups drew the same percentage of inferences "she will study late in the library". This example demonstrates the human capability to draw *non-monotonic* inferences: It is not the additional conditional *per se* but an additional conditional that may hint a reasoner at exceptions of the α-conditional leading to a suppressed inference.

There are basically two approaches how a cognitive scientist may try to model the different inferences humans draw in the suppression task by a logical system: S/he can develop a new logic or s/he can investigate what already existing logics may explain. Most researchers from psychology have followed the latter and have investigated specific systems, e.g., System P [Neves *et al.*, 2002; Pfeifer and Kleiter, 2005], weak completion semantics with Łukasiewicz-logic [Dietz *et al.*, 2012] or Kleene-logic [Stenning and Lambalgen, 2008]. Each logic has its own properties and, to the best of our knowledge the relationships between Łukasiewicz (or Kleene) to de-facto standards as Reiter's default logic, or inference with System P, System Z, or the successful approach of c-representations remain unknown. Analyzing the most prominent non-monotonic logics on such core-examples can shed light on these systems as well from a formal perspective.

The goal of this paper is to investigate if and how the different non-monotonic logics differ inferentially between the β- and the γ-problems and compare these inferences to human inferences. That is, we switch the usual *normative perspective* that the system defines the norm the human reasoner is tested against such that we use the human reasoner as norm the systems are tested against. The human reasoner as a norm is given by experimental results. In the following we focus on the results provided by [Byrne, 1989]. A logic that draws the same distinction as humans do is nearer to the human inference process, can be used as a cognitive model and is more cognitively-adequate. And, this can be a first step towards a better understanding of commonsense reasoning and about the human reasoning process in general. It can bring problems from Cognitive Science to Artificial Intelligence and enrich cognitive science modeling by formal techniques from knowledge representation and reasoning.

The rest of the paper is structured as follows: In the next section we introduce some technical preliminaries. Afterwards we introduce five important non-monotonic logics and investigate the inferences for the suppression task. An evaluation and discussion of different approaches with consequences about the associated knowledge bases conclude the article.

Preliminaries

For each conditional "if e then l" the 'e' is called the antecedent and 'l' is called the consequent of the conditional. [Stenning and Lambalgen, 2008] proposed two ways how a formalism can be evaluated: The so-called *conceptual cognitive-adequacy* and the so-called *inferential cognitive-adequacy* of a reasoning system. The first investigates if the formal representation of a reasoning system is similar to human mental representations. The second investigates if the inferences a reasoning system draws are similar/identical to the human inferences. Applied to our investigations, everything depends on how we can interpret the conditional (that would be the conceptual cognitive-adequacy) and the inference system that we apply (the inferential cognitive-adequacy). There are different ways how a conditional given in natural language as "if e then l" can be interpreted. The first possibility is to interpret it as material implication (e.g., propositional logic). We write shortly $e \rightarrow l$, whenever we refer to this monotonic interpretation. And there are at least two general ways of how we can extend this by introducing exceptions in a conditional statement. Exceptions can be represented in two ways: (1) In the *antecedent*, e.g., [Dietz *et al.*, 2012; Stenning and Lambalgen, 2008] for the weak completion semantics. In this case abnormality predicates (ab) capture the exception, i.e., for the example above we write $l \leftarrow e \wedge \neg \, \mathrm{ab}_1$. Stenning and Lambalgen call this a "licence for implication". And, these logics use a third truth value: the classical two values like $\top$ (for true) or $\bot$ (for false), and additionally, u, for unknown. (2) In *interpreting the implication*; e.g., Reiter's Default logic, System P, System Z, or c-representations. Instead of using the material implication, in commonsense reasoning a conditional represents a notion of plausibility. The classical consequence relation $\models$ with $e \models l$ represents: if e is true, then l *must* be true. In contrast the non-monotonic consequence relation [Kraus *et al.*, 1990] use $\mid\!\sim$ with $e \mid\!\sim l$ meaning: if e is true, then *typically* l is true as well. Applied to the suppression task $e \mid\!\sim l$ does not imply $e \wedge o \mid\!\sim l$. Most 'systems' try to characterize $\mid\!\sim$ by specific rules (see below).

Other than the material conditional with its monotonic interpretation we use conditionals $(B|A)$ to express rules that may have exceptions, "if A then *usually* B". $(B|A)$ is verified if both the antecedent and consequence are true and falsified if the antecedent is true and the consequent is false. If the antecedent is false, the evaluation of the conditional is *undefined* and the conditional is not applicable. A conditional is *tolerated* from a set of conditionals iff there is a world that verifies the conditional and does not falsify any conditional in the set [de Finetti, 1974; Kern-Isberner, 2001].

A *knowledge base* Δ in the following is a finite set of conditionals over a language. A knowledge base Δ is *consistent* iff for every nonempty subset $\Delta' \subseteq \Delta$ there is a conditional $(B|A) \in \Delta'$ that is tolerated by Δ'. This is equivalent to saying that there is a maximal (with respect to set inclusion) partitioning of the knowledge base $\Delta = \Delta_0 \uplus ... \uplus \Delta_m$ such that each conditional in a partition Δ_i $0 \leq i < m$ is tolerated by the set $\bigcup_{j=i+1}^{m} \Delta_j$ [Pearl, 1990]. For the β-case of the example above it is $\Delta^\beta = \{(l|e), (l|t), (e|\top)\}$. As humans may interpret the γ-conditional differently, we investigate the following three knowledge bases:

$$\Delta^\gamma = \{\delta_1 : (l|e), \delta_2 : (l|o), \delta_3 : (e|\top)\} \qquad (\gamma\text{-case})$$

$$\Delta^{\gamma'} = \{\delta_1 : (l|e), \delta_4 : (o|l), \delta_3 : (e|\top)\} \qquad (\gamma'\text{-case})$$

$$\Delta^{\gamma''} = \{\delta_1 : (l|e), \delta_5 : (\bar{l}|\bar{o}), \delta_3 : (e|\top)\} \qquad (\gamma''\text{-case})$$

Here, Δ^γ is the implementation of the γ-case from the original work of [Stenning and Lambalgen, 2008] with δ_2 being the literal meaning of the premise γ. We extend this with two additional implementations, γ' and γ'', where δ_4 encodes the conditional "if she is in the library, the library is/must be open" and δ_5 refers to: "if the library is not open, she is not/cannot be in the library". Both are two possible alternative interpretations humans may draw if they read/hear δ_2; others are as well possible but we focus exemplarically on these three.

Systems of nonmonotonic reasoning and the suppression task

In this section we briefly introduce some prominent nonmonotonic logics (for an overview of formalizations see Table 2). We test inferences regarding l of these reasoning systems for the γ, γ', and γ''-cases. If the logics infer l for these cases then they do for the β-case as well.

We start modeling the Suppression Task with the prominent and de facto standard of inference systems, System P, proceed with logic programming and Reiter's default logic and finally model the task in systems based on ordinal conditional functions (OCF).

1. System P

A de-facto standard in nonmonotonic reasoning is System P. It computes a preferential consequence relation $\mathrel{|\!\sim}$ based on the KLM-rules in the table below [Kraus *et al.*, 1990]:

Reflexivity	for all $A \in \mathfrak{L}$ it holds that $A \mathrel{|\!\sim} A$
Left Logical Equiv.	$A \equiv B$ and $B \mathrel{|\!\sim} C$ imply $A \mathrel{|\!\sim} C$
Right weakening	$B \models C$ and $A \mathrel{|\!\sim} B$ imply $A \mathrel{|\!\sim} C$
Cautious Monotony	$A \mathrel{|\!\sim} B$ and $A \mathrel{|\!\sim} C$ imply $AB \mathrel{|\!\sim} C$
CUT	$A \mathrel{|\!\sim} B$ and $AB \mathrel{|\!\sim} C$ imply $A \mathrel{|\!\sim} C$
OR	$A \mathrel{|\!\sim} C$ and $B \mathrel{|\!\sim} C$ imply $(A \vee B) \mathrel{|\!\sim} C$

Definition 1 [Goldszmidt and Pearl, 1991] *For a given preferential structure $(W, <)$ consisting of a set of worlds W and a partial order $<$ on the worlds, we define p-entailment ($\phi \mathrel{|\!\sim}^P \psi$) if ψ holds in all $<$-minimal models of ϕ.*

Whether a formula p-entails another from a knowledge base can be computed by checking whether the knowledge base can be extended consistently with the inverse of the conditional formed from both formulae, formally:

Proposition 1 [Goldszmidt and Pearl, 1996] *Let Δ be a knowledge base and A and B be formulas. Formula A p-entails B in the context of Δ iff $\Delta \cup \{(\overline{B}|A)\}$ is inconsistent.*

For all three cases, we check whether we can entail l or $\bar{l}$ from the knowledge base. We check whether Δ^γ, $\Delta^{\gamma'}$ or $\Delta^{\gamma''}$ are inconsistent with $(\bar{l}|\top)$ or $(l|\top)$, respectively.

In the γ-case, $\Delta^\gamma \cup \{(l|\top)\}$ is consistent, the world elo, for instance, verifies all conditionals. On the other hand, $\Delta^\gamma \cup \{(\bar{l}|\top)\}$ is inconsistent since there is no world that verifies at least one of the conditionals and accepts all others. Therefore, in the γ-case we have $\top \mathrel{|\!\sim}^P_{\Delta^\gamma} l$ and $\top \mathrel{|\!\not\sim}^P_{\Delta^\gamma} \bar{l}$, and hence 'she will study late in the library'.

In the γ'-case, $\Delta^{\gamma'} \cup \{(l|\top)\}$ is consistent, the world elo, for instance, verifies all conditionals. On the other hand, $\Delta^{\gamma'} \cup \{(\bar{l}|\top)\}$ is inconsistent since there is no world that verifies at least one of the conditionals and accepts all others. Therefore, in the γ'-case we have $\top \mathrel{|\!\sim}^P_{\Delta^{\gamma'}} l$ and $\top \mathrel{|\!\not\sim}^P_{\Delta^{\gamma'}} \bar{l}$, and hence 'she will study late in the library'.

In the γ''-case, $\Delta^{\gamma''} \cup \{(l|\top)\}$ is consistent, the knowledge base can be partitioned in the tolerance partitions $\Delta_0^{\gamma''} = \{\delta_1, \delta_3, (l|\top)\}$ and $\Delta_1^{\gamma''} = \{\delta_5\}$. On the other hand, $\Delta^{\gamma''} \cup \{(\bar{l}|\top)\}$ is inconsistent as there is no world that verifies at least one of the conditionals and accepts all others. Therefore, in the γ''-case we have $\top \mathrel{|\!\sim}^P_{\Delta^{\gamma''}} l$ and $\top \mathrel{|\!\not\sim}^P_{\Delta^{\gamma''}} \bar{l}$, and hence 'she will study late in the library'.

In System P we can infer for all three cases Δ^γ, $\Delta^{\gamma'}$, $\Delta^{\gamma''}$ that 'she will study late in the library'. As a consequence System P does not replicate the suppression effect.

2. Logic programming approaches

Logic programming with a weak completion semantics has been recently successfully applied to the γ-problems of the Suppression Task. As outlined above Stenning and van Lambalgen [Stenning and Lambalgen, 2008] and Dietz and colleagues [Dietz *et al.*, 2012] claim that conditionals should be encoded by "licenses for implications". For example, the conditional *if she has an essay to finish, she will study late in the library* or short ($l \leftarrow e$) should be encoded by the clause $l \leftarrow e \wedge \overline{\mathrm{ab}_1}$, where ab_1 is an *abnormality* predicate which expresses that l holds if e holds and nothing abnormal is known. In Table 1 two logic programs are presented for the two examples of the Suppression Task [Dietz *et al.*, 2012]. A second aspect is that despite the great success of two valued logics in artificial intelligence or cognitive science, the truth of a statement or premise cannot always be determined and so it might be cognitively plausible to introduce a third truth value, namely *unknown/undefined* [Łukasiewicz, 1920]. By introducing a third truth value, there are many possibilities for defining truth tables for the connectives [Kleene, 1952; Łukasiewicz, 1920; Fitting, 1985]. The Łukasiewicz logic has the model intersection property [Hölldobler and Kencana Ramli, 2009], i.e., the intersection of two models is a model. This property entails the existence of least models. The so-called weak-completion semantics process works as follows [Hölldobler and Kencana Ramli, 2009]:

1. Replace all clauses with the same head by a disjunction of the body elements, i.e., $A \leftarrow B_1, \ldots, A \leftarrow B_n$ by $A \leftarrow B_1 \vee \ldots \vee B_n$.
2. Replace all occurrences of $\leftarrow$ by $\leftrightarrow$.

The resulting set of equivalences is called the *weak completion* and the model intersection property holds for weakly completed programs [Hölldobler and Kencana Ramli, 2009] guaranteeing the existence of a least model.

The abnormality predicates (e.g., ab_1) represent abnormal cases: For instance, ab_1 is true when *the library does not stay open* and ab_3 is true when *she does not have an essay to finish*. The logic programs and the inferences can be found in Table 1. The results show that only in the γ-case the WCS does not draw the l-inference, but in the alternative cases γ' and γ'' it can be inferred. The cases γ' and γ'' require additional empirical data from human participants. Weak completion semantics is very sensitive to additional information.

Theorem 1 *Cautious monotony does not hold in the weak completion semantics.*

Table 1: The weak completion semantics approach for the β, γ, γ'-cases. γ'' is analog to γ. The ab represents abnormality predicates. Percentages present modus ponens drawn by the participants [Byrne, 1989]. Adapted from [Dietz et al., 2012].

Problems	β	γ	γ'
Program	$l \leftarrow e \wedge \overline{\mathrm{ab}_1}$ $l \leftarrow t \wedge \overline{\mathrm{ab}_2}$ $\mathrm{ab}_1 \leftarrow \bot$ $\mathrm{ab}_2 \leftarrow \bot$ $e \leftarrow \top$	$l \leftarrow e \wedge \overline{\mathrm{ab}_1}$ $l \leftarrow o \wedge \overline{\mathrm{ab}_3}$ $\mathrm{ab}_1 \leftarrow \overline{o}$ $\mathrm{ab}_3 \leftarrow \overline{e}$ $e \leftarrow \top$	$l \leftarrow e \wedge \overline{\mathrm{ab}_1}$ $o \leftarrow l \wedge \overline{\mathrm{ab}_4}$ $\mathrm{ab}_1 \leftarrow \overline{o}$ $\mathrm{ab}_4 \leftarrow \overline{e}$ $e \leftarrow \top$
wcs	$l \leftrightarrow (e \wedge \overline{\mathrm{ab}_1})$ $\vee (t \wedge \overline{\mathrm{ab}_2})$ $\mathrm{ab}_1 \leftrightarrow \bot$ $\mathrm{ab}_2 \leftrightarrow \bot$ $e \leftrightarrow \top$	$l \leftrightarrow (e \wedge \overline{\mathrm{ab}_1})$ $\vee (o \wedge \overline{\mathrm{ab}_3})$ $\mathrm{ab}_1 \leftrightarrow \overline{o}$ $\mathrm{ab}_3 \leftrightarrow \overline{e}$ $e \leftrightarrow \top$	$l \leftrightarrow (e \wedge \overline{\mathrm{ab}_1})$ $o \leftrightarrow (l \wedge \overline{\mathrm{ab}_4})$ $\mathrm{ab}_1 \leftrightarrow \overline{o}$ $\mathrm{ab}_4 \leftrightarrow \overline{e}$ $e \leftrightarrow \top$
Least Model	$(\{e, l\}, \{\mathrm{ab}_1, \mathrm{ab}_2\})$	$(\{e\}, \{\mathrm{ab}_3\})$	$(\{e\}, \{\mathrm{ab}_4\})$
Percentage of participants inferring l	96%	38%	n/a

Cautious monotony is defined by, if $\Pi \models a$, $\Pi \models b$ then $\Pi \cup \{a\} \models b$. To show that cautious monotony does not hold we adapt an idea of [Baral, 2003, p. 338]. Consider the following program. $\Pi = \{\neg b \leftarrow a, c \wedge \neg a \leftarrow b, a \leftarrow c\}$. Then we can infer by weak completion that a and c follow. For $\Pi \cup \{c\}$, however, we cannot conclude a anymore. So in this respect the weak completion semantics seems to be a very cautious approach in the class of non-monotonic logics.

3. Reiter's Default Logic

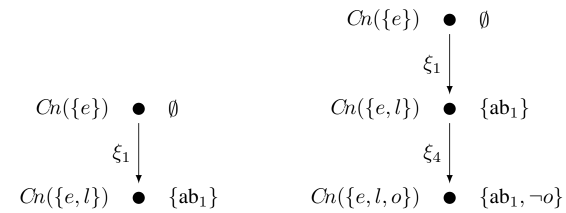
Reiter's default logic [Reiter, 1980] is based on a tuple (W, D) with a classical background theory W and a set of default rules $D = \{\xi_1, \xi_2, ..., \xi_n\}$ where each default $\xi = \frac{pre(\xi) : just(\xi)}{cons(\xi)}$ is composed of a precondition (the formula $pre(\xi)$), a set of justifications $just(\xi)$ and a set of consequences $cons(\xi)$. A default ξ is *applicable* to a deductively closed set $Cn(\mathcal{A})$ iff $pre(\xi) \in Cn(\mathcal{A})$ and $\neg B \notin Cn(\mathcal{A})$ for every $B \in just(\xi)$. A *(default) process* Π is a finite sequence of defaults $(\xi_{\Pi_1}, ..., \xi_{\Pi_m})$, $\delta_{\Pi_i} \in D$ for all $1 \leq i \leq m$ with the two sets $In(\Pi) = Cn(W \cup \{cons(\xi) | \xi \in \Pi)$ and $Out(\Pi) = \{\neg A | A \in just(\xi), \xi \in \Pi\}$ such that each default ξ is applicable to the In-set of the foregoing defaults. A process is *successful* iff $In(\Pi) \cap Out(\Pi) = \emptyset$ and *closed* iff every $\xi \in D$ that is applicable to $In(\Pi)$ is an element of Π. $\mathcal{E}$ is an *extension* of (W, D) iff there is a closed and successful process Π with $\mathcal{E} = In(\Pi)$ [Antoniou, 1997]. A formula is inferred credulously from (W, D) under Reiter's logic iff it is element of the union of all extensions and sceptically iff it is element of the intersection of all extensions $\mathcal{E}$ of (W, D).

For the suppression task $W = \{e\}$ is given. We transfer the implementation for Logic programming [Stenning and Lambalgen, 2008; Dietz *et al.*, 2012] as shown in the previous section to Reiter defaults and obtain the sets

$$D^{\gamma} = \left\{\xi_1 = \frac{e : \overline{\mathrm{ab}_1}}{l}; \xi_2 = \frac{\overline{o} : \mathrm{ab}_1}{l}; \xi_3 = \frac{o : \overline{\mathrm{ab}_2}}{l}; \xi_4 = \frac{\overline{e} : \mathrm{ab}_2}{\mathrm{ab}_2}\right\}$$

$$D^{\gamma''} = D^{\gamma'} = \left\{\xi_1 = \frac{e : \overline{\mathrm{ab}_1}}{l}; \xi_4 = \frac{l : o}{o}; \xi_2 = \frac{\overline{o} : \mathrm{ab}_1}{\mathrm{ab}_1}\right\}$$

Figure 1: The process trees visualising the processes for the suppression task in Reiter's Default Logic.



(a) The case D^{γ} (b) The cases $D^{\gamma'}$ and $D^{\gamma''}$

As shown in the process tree in Figure 1, in all cases there is only one extension, hence credulous and sceptical Reiter inference is identical for this task. And l is in every extension of the default logic, and we can credulously and sceptically infer for both tasks that given D^{γ}, $D^{\gamma'}$, and $D^{\gamma''}$ from the fact that 'she has an essay to write' we can infer that 'she will study late in the library', formally (Reiter^c represents the credulous case and Reiter the sceptical case):

$$e \mid\!\sim^{\mathrm{Reiter}^c}_{D^{\gamma}} l, \quad e \mid\!\sim^{\mathrm{Reiter}}_{D^{\gamma}} l, \quad e \mid\!\sim^{\mathrm{Reiter}^c}_{D^{\gamma'}} l, \quad \text{and} \quad e \mid\!\sim^{\mathrm{Reiter}}_{D^{\gamma'}} l.$$

4. OCF based systems

An *Ordinal Conditional Function* (OCF) [Spohn, 2012] is a function $\kappa : \Omega \rightarrow \mathbb{N}_0^{\infty}$ assigning to each world $\omega \in \Omega$ an implausibility rank, that is, the higher $\kappa(\omega)$, the less plausible the world is. Moreover, as a normalization condition, there must be at least one world ω such that $\kappa(\omega) = 0$. The rank of a formula $A \in \mathfrak{L}$ is the minimal rank of all worlds that satisfy A, $\kappa(A) = \min\{\kappa(\omega) | \omega \models A\}$. The rank of a conditional $(B|A)$ is the rank of the conjunction of premise and conclusion normalised by the rank of the premise, $\kappa(B|A) = \kappa(AB) - \kappa(A)$. κ *accepts* a conditional $(B|A)$ (written $\kappa \models (B|A)$) iff its verification is more plausible than its falsification. The inference relation of OCF is defined using preferential models [Makinson, 1994], defining that B is inferred from A using the OCF κ, formally $A \mid\!\sim_{\kappa} B$, iff the verification of the conditional $(B|A)$ is more plausible than its falsification, iff κ accepts the conditional, formally

$$A \mid\!\sim_{\kappa} B \quad \text{iff} \quad \kappa(AB) < \kappa(A\overline{B}) \quad \text{iff} \quad \kappa \models (B|A)$$

An OCF can be obtained inductively from a knowledge base. To model the suppression task we use System Z and c-representations for inductive reasoning.

4a. System Z

System Z uses the partitioning of the consistency test of a knowledge base from the preliminaries as a notion of exceptionality. It rates the implausibility of each world by the most exceptional conditional falsified, that is, System Z assigns to each world the maximum partition number which contains conditionals falsified by this world (incremented by 1 because the numbering of partitions is zero-based).

In the γ-case, all conditionals in the knowledge base are put into Δ_0^{γ}. For example, the world elo verifies every conditional in the knowledge base, hence every conditional is tolerated by Δ_0^{γ}. Therefore we get the ranking function which is

Table 2: Typical comprehension of the natural-language conditionals of the suppression task in the considered logics.

Natural Language	Conditional			Inference				
	material	indicative	weak completion	cred. Reiter	Reiter	System P	System Z	c-rep.
If she has an (e)ssay to write then she will study late in the (l)ibrary.	$e \to l$	$(l\|e)$	$l \leftarrow e \wedge \neg \mathrm{ab}_1$	$e \mid\!\sim^{\mathrm{Reiter}^c}_{\Delta} l$	$e \mid\!\sim^{\mathrm{Reiter}}_{\Delta} l$	$e \mid\!\sim^{P}_{\Delta} l$	$e \mid\!\sim^{Z}_{\Delta} l$	$e \mid\!\sim^{c}_{\Delta} l$
If the library is (o)pen then she will study late in the (l)ibrary.	$o \to l$	$(l\|o)$	$l \leftarrow o \wedge \neg \mathrm{ab}_2$	$o \mid\!\sim^{\mathrm{Reiter}^c}_{\Delta} l$	$o \mid\!\sim^{\mathrm{Reiter}}_{\Delta} l$	$o \mid\!\sim^{P}_{\Delta} l$	$o \mid\!\sim^{Z}_{\Delta} l$	$o \mid\!\sim^{c}_{\Delta} l$
If she has a (t)extbook to read then she will study late in the (l)ibrary.	$t \to l$	$(l\|t)$	$l \leftarrow t \wedge \neg \mathrm{ab}_3$	$t \mid\!\sim^{\mathrm{Reiter}^c}_{\Delta} l$	$t \mid\!\sim^{\mathrm{Reiter}}_{\Delta} l$	$t \mid\!\sim^{P}_{\Delta} l$	$t \mid\!\sim^{Z}_{\Delta} l$	$t \mid\!\sim^{c}_{\Delta} l$
She has an (e)ssay to write.	e	$(e\|\top)$	$e \leftarrow \top$	$\top \mid\!\sim^{\mathrm{Reiter}^c}_{\Delta} e$	$\top \mid\!\sim^{\mathrm{Reiter}}_{\Delta} e$	$\top \mid\!\sim^{P}_{\Delta} e$	$\top \mid\!\sim^{Z}_{\Delta} e$	$\top \mid\!\sim^{c}_{\Delta} e$

Table 3: OCFs obtained by System Z and minimal c-representations for the cases γ, γ', and γ''.

ω	$e\,l\,o$	$e\,l\,\bar{o}$	$e\,\bar{l}\,o$	$e\,\bar{l}\,\bar{o}$	$\bar{e}\,l\,o$	$\bar{e}\,l\,\bar{o}$	$\bar{e}\,\bar{l}\,o$	$\bar{e}\,\bar{l}\,\bar{o}$
$\kappa^Z_{\Delta^\gamma}(\omega)$	0	0	1	1	1	1	1	1
$\kappa^c_{\Delta^\gamma}(\omega)$	0	1	1	1	1	1	2	1
$\kappa^Z_{\Delta^{\gamma'}}(\omega)$	0	1	1	1	1	1	1	1
$\kappa^c_{\Delta^{\gamma'}}(\omega)$	0	1	2	1	1	2	1	1
$\kappa^Z_{\Delta^{\gamma''}}(\omega)$	0	2	1	1	1	2	1	1
$\kappa^c_{\Delta^{\gamma''}}(\omega)$	0	2	1	1	1	3	1	1

shown in Table 3. For System Z we have $\kappa^Z_{\Delta^\gamma}(l) = 0 < 1 = \kappa^Z_{\Delta^\gamma}(\bar{l})$, hence it can be derived that 'she will study late in the library'.

In the γ'-case, we again have $\Delta_0^{\gamma'} = \{\delta_1, \delta_3, \delta_4\}$ because, e.g., the world elo verifies every conditional in the knowledge base which gives the OCF (Table 3): $\kappa^Z_{\Delta^{\gamma'}}(l) = 0 < 1 = \kappa^Z_{\Delta^{\gamma'}}(\bar{l})$ hence 'she will study late in the library'.

In the γ''-case, we have the partitions $\Delta_0^{\gamma''} = \{\delta_1, \delta_3\}$ and $\Delta_1^{\gamma''} = \delta_5$. The OCF (Table 3) is again $\kappa^Z_{\Delta^{\gamma''}}(l) = 0 < 1 = \kappa^Z_{\Delta^{\gamma''}}(\bar{l})$ hence 'she will study late in the library'. So the System Z does not show the suppression effect.

4b. c-representations

System Z combines conditionals in the knowledge base by their exceptionality. It is known that this approach may lead to neglecting the effect of conditionals (the so-called "Drowning Problem" [Pearl, 1990; Benferhat *et al.*, 1993]). Other than that, the approach of c-representations [Kern-Isberner, 2001; 2004] assigns to each conditional an individual impact $\kappa_i^- \in \mathbb{N}_0$ as abstract weight to each conditional in the knowledge base $\mathcal{R} = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$. The rank of a world is the combined impact of all falsified conditionals, so a c-representation κ^c_Δ is an OCF defined by

$$\kappa^c_\Delta(\omega) = \sum_{\omega \models \phi_i \wedge \neg \psi_i} \kappa_i^-, \qquad (1)$$

where the individual impacts $\kappa_i^- \in \mathbb{N}_0$ are chosen such that κ^c_Δ is admissible with respect to Δ, which is the case if the impacts satisfy the following system of inequations [Kern-Isberner, 2001; 2004]:

$$\kappa_i^- > \min_{\substack{\omega \in \Omega \\ \omega \models \phi_i \wedge \psi_i}} \Big\{ \sum_{\substack{1 \leq j \leq n, j \neq i \\ \omega \models \phi_j \wedge \neg \psi_j}} \kappa_j^- \Big\} - \min_{\substack{\omega \in \Omega \\ \omega \models \phi_i \wedge \neg \psi_i}} \Big\{ \sum_{\substack{1 \leq j \leq n, j \neq i \\ \omega \models \phi_j \wedge \neg \psi_j}} \kappa_j^- \Big\} \qquad (2)$$

Table 4: System Z and c-representations mimicking the weak completion semantics approach for the β, γ, γ'-cases.

Problems	β	γ	γ'
knowledge base	$(l\|e)$ $(l\|t)$	$(l\|eo)$	$(l\|eo)$ $(o\|le)$
Belief sets	$Cn(el)$	$Cn(el \vee e\bar{l}\bar{o})$	$Cn(elo \vee e\bar{l}\bar{o})$
wcs knowledge base	$(l\|e)$ $(l\|t)$ $(e \vee t\|l)$	$(l\|eo)$ $(eo\|l)$	$(l\|eo)$ $(eo\|l)$ $(o\|le)$ $(le\|o)$
Belief sets	$Cn(el)$	$Cn(elo \vee e\bar{l}\bar{o})$	$Cn(elo \vee e\bar{l}\bar{o})$
Percentage	96%	38%	n/a

Applying c-representations to the γ-case the system of inequations (2) can be solved minimally with the values $\kappa_1^- = 1$, $\kappa_2^- = 0$ and $\kappa_3^- = 1$ which gives us the OCF $\kappa^c_{\Delta^\gamma}(\omega)$ in Table 3. We have $\kappa^c_{\Delta^\gamma}(l) = 0 < 1 = \kappa^c_{\Delta^\gamma}(\bar{l})$ hence 'she will study late in the library'.

In the γ'-case, the system of inequations (2) can be solved with the values $\kappa_1^- = 1$, $\kappa_4^- = 1$ and $\kappa_3^- = 1$ which gives us the OCF $\kappa^c_{\Delta^{\gamma'}}(\omega)$ in Table 3. We have $\kappa^c_{\Delta^{\gamma'}}(l) = 0 < 1 = \kappa^c_{\Delta^{\gamma'}}(\bar{l})$ hence 'she will study late in the library'.

In the γ''-case, the system of inequations (2) can be solved with the values $\kappa_1^- = 1$, $\kappa_5^- = 1$ and $\kappa_3^- = 1$ hence the OCF is $\kappa^c_{\Delta^{\gamma''}}(\omega)$ (cp. Table 3). We have $\kappa^c_{\Delta^{\gamma''}}(l) = 0 < 1 = \kappa^c_{\Delta^{\gamma''}}(\bar{l})$ hence 'she will study late in the library'. A possibility is to reformulate the knowledge base as we can see in the next subsection.

4c. Mimicking weak completion semantics (WCS)

The weak completion semantics is able to make a difference between the cases β and γ resp. γ'. However, as can be seen from Table 1, there is a slight difference already in modelling these cases – the abnormality predicates in cases γ resp. γ' are linked to $\bar{e}$ and $\bar{o}$ which is not done for case β. We show that we can achieve at least very similar effects to the weak completion semantics by using System Z or c-representations when building them from conditional knowledge bases which mimick the modellings from Table 1. We consider both the modelling obtained from the logic program and the modelling induced by WCS. Instead of putting the evidential information e into the knowledge base, we compute the final results by conditioning the κ-function obtained from the generic knowledge base by System Z resp. c-representations on e because this simulates better the distinction between generic and evidential knowledge. However technically, in the considered cases, putting e into the knowledge base right from the beginning would not make a significant difference. The computations are very similar to the ones performed in the previous subsections: we summarize the results in Table 4.

In all cases, the most plausible beliefs are the same for System Z and c-representations (there are slight differences in the resulting κ-functions), so we do not distinguish between these two approaches in the table. In the end, we find that in the β-case, the agent believes el, but in the γ and γ'-cases, the agent believes e while no longer being certain about l. Interestingly, this effect is already obtained from modelling the logic program (without WCS). It is only in the γ-case, that there is a minor difference at all between the belief sets of the knowledge bases with or without weak completion semantics: Here, WCS induces a stronger equivalence between l and o.

Conclusion and Discussion

A core problem in AI, many non-monotonic logics have been evaluated on, is 'Tweety' the famous commonsense reasoning problem about a penguin that is a spoiler for the often accepted statement that birds can fly. A successful AI reasoning system must handle a class/subclass inheritance problem. Another feature relevant in commonsense reasoning among humans is *cautious reasoning*. Hence, we investigated the suppression task, a core research paradigm in Cognitive Science/Psychology, that shows how humans deal in the light of new information. We evaluated the γ-case (and variations) to test, if humans may have interpreted the second premises differently. Our motivation was two-fold to (1) test the inferences from the different non-monotonic logics for a core problem relevant in cognitive science and, as a subsequent goal, (2) to learn more about the relationships between the logics. We analyzed many of the de-facto standards in NML, like Reiter's default logic, System P, System Z, and c-representations with human inferences, how they deal with such cautious reasoning problems. To test the inferences (1) we could show that the weak completion semantics, c-representations and System Z make different inferences for the β and for the γ, γ', and γ''-cases. The other systems Reiter's default logic and System P made no distinction for the three investigated representations γ, γ', and γ''. However, there could be additional alternative interpretations of the first and second premise. Especially the interpretation of the conclusion of the premise α (Table 1), that is, whether she will stay late in the library, is intricate. By separating this sentence into "staying / working late" and "being in the library" α can be interpreted to be "if she has an essay to write and she stays / works late, then she will be in the library" as well as to be "if she has an essay to write, then she will stay / work late and be in the library". By separating the knowledge in this way, it is possible to trigger hidden background knowledge that (usually) libraries are not accessible 24-hours a day, so "staying late" may trigger an exception to the library being open, if the participant is made aware that it could be otherwise by γ. By using this formalisation, c-revision, i.e., belief revision with c-representations [Kern-Isberner, 2001] captures the suppression effect and makes as well a difference between the β and γ-cases. The basic idea is that we need to extend the notion of a knowledge base by a temporal order of the information processed by humans. Especially the conceptual-adequacy needs to be further investigated from a formal and cognitive perspective. In the last decade System P has been regarded as a possible candidate for capturing human reasoning [Pfeifer and Kleiter, 2005; Neves *et al.*, 2002]. But, previous empirical research showed that none of the three systems C, CL, and P could be shown to be cognitively-adequate [Kuhnmünch and Ragni, 2014], i.e., some inferences are rarely drawn. Additionally, System P does not make any distinction between the β and γ-case, so further research is necessary to test if System P captures human reasoning processes. The formal analysis inspires future empirical research questions and we need to thoroughly investigate the human interpretation of the premises. Regarding the second point: As the Tweety-problem shows differences between the nonmonotonic systems, the suppression task does as well. Although the weak-completion-semantics have not yet, to the best of our knowledge, been formally compared to system-based non-monotonic logics, we derive:

Theorem 2 *Weak completion semantics is not contained in Reiter's default logic, System P, Z, and c-representations.*

The first two follow from the difference on the suppression task. And the latter two follow from Theorem 1 and from the fact that for System Z and c-representation holds the even stronger principle of rational monotony.

The motivation of this article was to investigate to which extend de-facto standard logics are able to simulate human inference processes and could serve potentially as cognitive models. Some non-monotonic logics seem to be adequate to describe human commonsense reasoning. The investigated de-facto logics have many interesting properties that inspire further empirical research on human reasoning processes, its possible nonmonotonic properties, and allow to develop human-oriented commonsense reasoning AI-systems.

Acknowledgments

This work is supported by DFG-Grants KI1413/5-1 to G. Kern-Isberner and by RA1934 2/1 and a Heisenberg DFG fellowship RA1934 3/1 to M. Ragni. C. Eichhorn is supported by Grant KI1413/5-1.

References

[Antoniou, 1997] G. Antoniou. *Nonmonotonic reasoning*. MIT Press, Cambridge, Mass., 1997.

[Baral, 2003] C. Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.

[Benferhat *et al.*, 1993] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency Management and Prioritized Syntax-Based Entailment. In *Proc. of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI'93)*, volume 1, pages 640–647, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers.

[Byrne, 1989] R. M. Byrne. Suppressing valid inferences with conditionals. *Cognition*, 31:61–83, 1989.

[Byrne, 2007] R.M.J. Byrne. *The Rational Imagination: How People Create Alternatives to Reality*. MIT Press, 2007.

[de Finetti, 1974] B. de Finetti. *Theory of Probability*, volume 1,2. John Wiley and Sons, New York, NY, USA, 1974.

[Dietz *et al.*, 2012] E. Dietz, S. Hölldobler, and M. Ragni. A Computational Approach to the Suppression Task. In N. Miyake, D. Peebles, and R.P. Cooper, editors, *Proc. of the 34th Annual Conference of the Cognitive Science Society*, pages 1500–1505, Austin, TX, 2012. Cognitive Science Society.

[Fitting, 1985] M. Fitting. A Kripke-Kleene Semantics for Logic Programs. *Journal of Logic Programming*, 2(4):295–312, 1985.

[Goldszmidt and Pearl, 1991] M. Goldszmidt and J. Pearl. On the consistency of defeasible databases. *Artificial Intelligence*, 52(2):121 – 149, 1991.

[Goldszmidt and Pearl, 1996] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1-2):57–112, 1996.

[Hölldobler and Kencana Ramli, 2009] S. Hölldobler and C. D. Kencana Ramli. Logic programs under three-valued Łukasiewicz semantics. In P. M. Hill and D. S. Warren, editors, *Logic Programming, 25th International Conference, ICLP 2009*, volume 5649 of *LNCS*, pages 464–478, Heidelberg, 2009. Springer.

[Kern-Isberner, 2001] G. Kern-Isberner. *Conditionals in Nonmonotonic Reasoning and Belief Revision*. Number 2087 in LNCS. Springer, Berlin, DE, 2001.

[Kern-Isberner, 2004] G. Kern-Isberner. A thorough axiomatization of a principle of conditional preservation in belief revision. *Annals of Mathematics and Artificial Intelligence*, 40:127–164, 2004.

[Kleene, 1952] Stephen Cole Kleene. *Introduction to metamathematics*. Bibl. Matematica. North-Holland, Amsterdam, 1952.

[Kraus *et al.*, 1990] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence Journal*, 44:167–207, 1990.

[Kuhnmünch and Ragni, 2014] G. Kuhnmünch and M. Ragni. Can Formal Non-monotonic Systems Properly Describe Human Reasoning? In P. Bello, M. Guarini, M. McShane, and B. Scassellati, editors, *Proc. of the 36th Annual Conference of the Cognitive Science Society*, pages 1222–1228. Austin, TX: Cognitive Science Society, 2014.

[Łukasiewicz, 1920] J. Łukasiewicz. O logice trójwartościowej. *Ruch Filozoficzny*, 5:169–171, 1920. English translation: On three-valued logic. In: Łukasiewicz J. and Borkowski L. (ed.). (1990). *Selected Works*, Amsterdam: North Holland, pp. 87–88.

[Makinson, 1994] D. Makinson. General Patterns in Nonmonotonic Reasoning. In Dov M. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 35–110. Oxford University Press, New York, NY, USA, 1994.

[Manktelow, 1999] K. I. Manktelow. *Reasoning and Thinking*. Psychology Press, Hove, UK, 1999.

[Neves *et al.*, 2002] R. Neves, J. Bonnefon, and E. Raufaste. An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):107–130, 2002.

[Oaksford and Chater, 2007] M. Oaksford and N. Chater. *Bayesian rationality: The probabilistic approach to human reasoning*. University Press, Oxford, 2007.

[Pearl, 1990] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh, editor, *Proc. of the 3rd conference on Theoretical aspects of reasoning about knowledge (TARK1990)*, pages 121–135, San Francisco, CA, USA, 1990. Morgan Kaufmann Publishers Inc.

[Pfeifer and Kleiter, 2005] N. Pfeifer and G. D. Kleiter. Coherence and nonmonotonicity in human reasoning. *Synthese*, 146(1-2):93–109, 2005.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1–2):81–132, 1980.

[Spohn, 2012] W. Spohn. *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford University Press, Oxford, UK, 2012.

[Stenning and Lambalgen, 2008] K. Stenning and M. Lambalgen. *Human reasoning and cognitive science*. Bradford Books. MIT Press, Cambridge, MA, 2008.

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