Minds & Machines (2017) 27:79–117
DOI 10.1007/s11023-016-9414-1

Formal Nonmonotonic Theories and Properties of Human Defeasible Reasoning

Marco Ragni[1] · **Christian Eichhorn**[2] · **Tanja Bock**[2] · **Gabriele Kern-Isberner**[2] · **Alice Ping Ping Tse**[1]

Received: 23 June 2016 / Accepted: 20 December 2016 / Published online: 23 February 2017

Abstract The knowledge representation and reasoning of both humans and artificial systems often involves conditionals. A conditional connects a consequence which holds given a precondition. It can be easily recognized in natural languages with certain key words, like "if" in English. A vast amount of literature in both fields, both artificial intelligence and psychology, deals with the questions of how such conditionals can be best represented and how these conditionals can model human reasoning. On the other hand, findings in the psychology of reasoning, such as those in the Suppression Task, have led to a paradigm shift from the monotonicity assumptions in human inferences towards nonmonotonic reasoning. Nonmonotonic reasoning is sensitive for information change, that is, inferences are drawn cautiously such that retraction of previous information is not required with the addition of new information. While many formalisms of nonmonotonic reasoning have been proposed in the field of Artificial Intelligence, their capability to model properties of human reasoning

✉ Marco Ragni
ragni@informatik.uni-freiburg.de

Christian Eichhorn
christian.eichhorn@tu-dortmund.de

Tanja Bock
tanja.bock@tu-dortmund.de

Gabriele Kern-Isberner
gabriele.kern-isberner@cs.uni-dortmund.de

Alice Ping Ping Tse
tse@informatik.uni-freiburg.de

[1] Research Group on the Foundations of Artificial Intelligence, Institut für Informatik, Technische Fakultät, Albert-Ludwigs-Universität Freiburg, Georges-Köhler-Allee, Geb. 052, 79110 Freiburg, Germany

[2] Technische Universität Dortmund, Department of Computer Science, Chair I - Information Engineering, 44221 Dortmund, Germany

has not yet been extensively investigated. In this paper, we analyzed systematically from both a formal and an empirical perspective the power of formal nonmonotonic systems to model (i) possible explicit defeaters, as in the Suppression Task, and (ii) more implicit conditional rules that trigger nonmonotonic reasoning by the keywords in such rules. The results indicated that the classical evaluation for the correctness of inferences has to be extended in the three major aspects (i) regarding the inference system, (ii) the knowledge base, and (iii) possible assumed exceptions for the rule.

Keywords Defeasible reasoning · Nonmonotonic logic · Suppression task · Cognitive modeling · Reasoning · Human reasoning · Knowledge representation · Cognitive systems

1 Introduction: Defeasible Reasoning in an Uncertain World

In everyday life conditional statements, such as "if Lisa (a student) has an essay to write, then she will study late in the library", play a central role in describing, for instance, rules, regularities, and causal chains. A conditional consists of an *antecedent* ("Lisa has an essay to write"), which acts as a precondition, and a *consequence* ("she will study late in the library") which holds in a classical logical sense given the antecedent is satisfied. Behavioural studies have shown that human reasoning does not, in general, follow the classical formal logical *material implication* inference rules but is far more complex than that (cf. Wason 1968; Johnson-Laird and Byrne 2002; Oaksford and Chater 2007; Klauer et al. 2007). Humans do not always endorse the consequence despite a satisfied antecedent (also known as the *modus ponens*, abbreviated "MP"). Nor do humans always derive from a negated consequence that the antecedent is false (also known as the *modus tollens*, abbreviated "MT"). In the following section, we briefly present the human reasoning process for possible defeating information with the Suppression Task (Byrne 1989).

1.1 Motivation and Behavioral Results: The Suppression Task (Byrne 1989)

Consider that a reasoner α hears that "If Lisa has an essay to write, then she will study late in the library" (abbreviated as: $e \rightarrow l$). Now, the reasoner α learns that Lisa does have an essay to write (e). Hence, α infers (like 95% of the participants in the Study in Byrne 1989) that "Lisa will study late in the library". Let us consider a second reasoner β who, in addition to what α knows (i.e., $e \rightarrow l$), learns the additional conditional, "If Lisa has a textbook to read, then she will study late in the library" ($t \rightarrow l$). By learning that Lisa has an essay to write (e), β infers that Lisa will study late in the library (again like 98% of the participants in the aforementioned study). Another reasoner γ, learns what α knows (i.e., $e \rightarrow l$) and the additional conditional that "If the library stays open, then Lisa will study late in the library". It is more possible that reasoner γ (than reasoners α and β) despite learning that "Lisa has an essay to write", γ does *not* endorse the conclusion that "Lisa will study late in the library" (nor did 62% of the participants in the Study in

Byrne 1989). Human reasoning process is sensitive to new information as inferences that would have been drawn before are not drawn now or only cautiously. This is a feature of *nonmonotonic reasoning*, allowing the retraction of previously valid inferences with newly gained information (Byrne 1989; Da Silva Neves et al. 2002). Classical logical reasoning (such as propositional logic) is *monotonic*, i.e., new information cannot lead to the retraction of previous inferences. We will briefly outline how the above mentioned finding can be modeled in major AI approaches to nonmonotonic reasoning in the next section.

1.2 Modeling the Suppression Task

The first cognitive model for the Suppression Task has been proposed by Byrne (1989). She suggested to model the difference between the β- and γ-cases by the difference in the understanding of the conditional (Byrne 1991; Byrne et al. 1999). The premises in logical formulation of the former case can be represented by $(e \vee t) \rightarrow l$ ("If Lisa has an essay to write *or* a textbook to read, then she will study late in the library"); while for the γ-case, it can be represented by the formula $(e \wedge o) \rightarrow l$ ("If she has an essay to write *and* the library is open, Lisa will study late in the library"). These two different formulations can lead to two different mental models. For the former case, application of modus ponens to derive l is allowed when e is given. However, in the latter, both e and o have to be given for the endorsement of l. In other words, the actual reasoning process depends on the underlying mental representations that are built from different logical representations/formulations of the integrated premise information.

Classical propositional logic (with the satisfaction relation $\models$) cannot model this observation. From $e \models l$ and e, l can be derived regardless of an additional information, namely $t \models l$ or $o \models l$. An idea which is in-between a formal and a cognitive model approach was proposed by Stenning and Lambalgen (2008). It was based on the idea to formulate, infinitive form re-represent a classical conditional as a "license for implication" and formulated the conditional as $l \leftarrow e \wedge \neg ab$ ("Lisa will study late in the library if she has an essay to write and the situation is not abnormal."). That means an abnormality predicates (*ab*) is inserted/introduced in the *antecedent*, to capture the exception. Additionally, a third truth value u (for *unknown*) is introduced besides the two classical values of $\top$ (for *true*) and $\bot$ (for *false*). By applying the Weak Completion Semantics (see below), a minimal model is built. The models for the β and γ cases above are not the same (e.g., Stenning and Lambalgen 2008; Dietz et al. 2012).

Recently a dispute about the expressive power of logic programs to explain human conditional reasoning has arisen (e.g., Oaksford and Chater 2016). With a formally inspired reasoning system, such as logic programming, one is able to reproduce the effects of the Suppression Task. However, there are many more formal systems in knowledge representation and reasoning (Antoniou 1997) which can deal with nonmonotonic reasoning, and some of them might intuitively be closer to human reasoning than others. A first comparison of the theories for the modus ponens (Ragni et al. 2016) has already shown that not all nonmonotonic systems are suitable for modeling human reasoning. Although some researchers have considered System P as the cognitive counterpart of human nonmonotonic reasoning (e.g., Bonnefon et al. 2006; Pfeifer and Kleiter 2005), it is not suitable for simulating human inferences for the Suppression Task (Ragni et al. 2016). In fact, it is even possible to prove that no suppression effect is possible in System P at all. Furthermore, only some of the outputs produced by System P are drawn by the majority of human reasoners, if several rules of System P are tested independently (Kuhnmünch and Ragni 2014). It is important to note that not only is the choice of the inference mechanisms relevant, but also the representation of the information.

1.3 Knowledge Representation, Conditionals, and Nonmonotonic Inferences

Strict and defeasible knowledge

Our first distinction is made regarding the type of knowledge. Nonmonotonic reasoning is reasoning about possible exceptions. Hence, we introduce the notion of *strict knowledge*, i.e., exceptions are not allowed, neither implicit nor explicit. We refer to any knowledge that allows for exceptions as *defeasible knowledge*. In the latter case, pieces of information can be considered as *supporters* or *defeaters* for the respective defeasible knowledge. We call the process of transforming strict knowledge to defeasible knowledge *weakening*. The strict and defeasible knowledge known to the reasoner is represented as the reasoner's *knowledge base*.

Conditionals and their representation

For some researchers, nonomotonicity in human reasoning, such as in the Suppression Task, has led to a so-called paradigm shift towards a probabilistic understanding of conditionals (Oaksford and Chater 2007). Proponents of such theories consider that a conditional, "if ϕ then ψ", is best represented by a conditional probability, $P(\psi \mid \phi)$. A study has demonstrated that the role of the relation between antecedent and consequence cannot be underestimated (Skovgaard-Olsen et al. 2016). A remaining problem is that most probabilistic accounts are not processing models, but they model the pattern of the answer distribution of the answer patterns instead. Probabilistic representation is not the only way in which such conditionals can be represented. For example, nonmonotonic systems are used in AI, especially in the area of knowledge representation and reasoning. A rational agent often has to be able to apply commonsense reasoning in order to reason with defaults (Reiter 1980; Beierle and Kern-Isberner 2014). Default reasoning already assumes that possible defeaters may exist. Hence, systems like Reiter default logic (Reiter 1980) deal with inferences based on assertions employing typicality knowledge, e.g., "birds typically fly". Commonsense reasoning plays an important role in formal argumentation systems, too (cf. Dung 1995; García and Simari 2004). The Suppression Task demonstrates that additional information can inhibit reasoners from drawing a classical MP or MT inference, by drawing reasoner's attention to possible defeaters for the conditional, for example, the necessary condition that a library is open so that Lisa can study late there. What is indirectly hidden is that participants may shift their processing of a classical conditional in the strict sense to a conditional that represents exceptions. An example of how

Table 1 Introducing defeasibility in a modus ponens like inference scheme by using the keyword "normally"

Type/premise	Classical cond	Cond weakening	Fact weakening	Cond fact weakening
Premise I	If Lisa has an essay to write, then she will study late in the library	If Lisa has an essay to write, then she will *normally* study late in the library	If Lisa has an essay to write, then she will study late in the library	If Lisa has an essay to write, then she will *normally* study late in the library
Premise II	Lisa will study late in the library	Lisa will study late in the library	*Normally*, Lisa will study late in the library	*Normally*, Lisa will study late in the library
Question	Does Lisa have an essay to finish?	Does Lisa have an essay to finish?	Does Lisa have an essay to finish?	Does Lisa have an essay to finish?

defeasibility can be introduced by the insertion of keywords such as "normally" is presented in Table 1. An intuitive prediction is that a reasoner should be more cautious in the case of weakened conditions, i.e., in the cases where defeaters are known or triggered by keywords that may hint at possible exceptions. Politzer (2005), for instance, investigated how human inferences change if they receive the classical premises but with a preceding sentence like: "it is not certain that ...?". He showed that this insertion of defeasible knowledge (uncertainty) can trigger comparable responses to the classical Suppression Task. The wording "not certain" does allow for a whole range of possible interpretations. In this study, we investigate how the insertion of a nonmonotonic keyword such as "normally" in conditionals can trigger defeasible reasoning. This article aims to compare de-facto standards in the field of nonmonotonic logics, for concrete problems, such as the Suppression Task (with explicit defeaters), and problems with weakening of the conditionals or/and facts with keywords (as implicit defeaters). We further tested the open question how well they can mimic specifics of human inferences.

Inference operator

We follow the idea of Gabbay (1985) and Makinson (1994) who argued to "bring some order into a rather chaotic field by considering the *output* of nonmonotononic systems". But in order to model the inferences we have to consider properties of the inference relation, $\mid\!\sim$, used by human reasoners. An inference is said to be based on a reasoning system Ψ that allows to decide whether a propositional formula ψ can be nonmonotonically derived from a propositional formula ϕ in the context of Ψ, written as $\phi \mid\!\sim^{\Psi} \psi$. And following (Makinson 1994), an inference operation $C(\cdot)$ defines the set of derivatives from a formula ϕ such that $\psi \in C_{\Psi}(\phi)$ if and only if $\phi \mid\!\sim^{\Psi} \psi$, i.e., $C_{\Psi}(\phi) = \{\psi \mid \phi \mid\!\sim^{\Psi} \psi\}$. These reasoning systems Ψ will be specified in later sections. The connection between the premise and the conclusion of a conditional $(\psi \mid \phi)$ that is accepted in Ψ holds, if and only if ϕ can be nonmonotonically derived from ψ by the system Ψ, formally $\phi \mid\!\sim^{\Psi} \psi$. We formally define this connection and inference with the concrete semantics in Sect. 3.

1.4 Research Questions

We briefly outline the research questions which arose during our previous discussion of the Suppression Task and the task of reasoning in a world with potential defeaters. This requires a thorough formal analysis and an empirical benchmarking for the two logically sound inference schemes modus ponens and modus tollens, on which we will focus in the following research questions:

RQ1: Can formal and de-facto standard nonmonotonic inference systems mimic the patterns we identified in human defeasible reasoning?

RQ2: What background information has to be additionally contained on the knowledge side (i.e., to be in the knowledge base) so that formal nonmonotonic inference systems show similar patterns?

RQ3: Are there differences between the systems for the two different forms of defeasible reasoning: (i) by introducing defeaters explicitly or (ii) by hinting implicitly at the existence of defeaters (i.e., insertion of the keyword "normally" in this study)?

The remainder of the paper is structured as follows: In Sect. 2 we briefly introduce propositional logic, conditionals, and various knowledge bases with an application of the aforementioned defeasible conditionals. Section 3 explains the de-facto standards in the field of nonmonotonic reasoning that are classified and tested with respect to their predictions. Section 4 applies theses approaches to the Suppression Task, which is implemented in different ways to show how differences in knowledge bases (strict or defeasible) may lead to different inferences by the systems. This section also summarizes findings in Ragni et al. (2016) and extends them by implementations of the MT case of the Suppression Task. Having demonstrated how formal differences in the implementation of the knowledge can lead to the differences in the inferences drawn, in Sect. 5, we systematically vary the strictness of belief, from strict knowledge without exceptions up to a knowledge base in which exceptions are allowed in both the rules and facts. In Sect. 6, we present three experiments to test the systems' predictions and human inferences with defeasible and strict knowledge. Finally, the predictions made by different systems are evaluated by multinomial process trees. We conclude the article with a summary and a general discussion in Sect. 7.

2 Preliminaries: Propositional Logic, Conditionals, and Knowledge Bases

Propositional logic is defined based on a set of variables, $\Sigma = \{V_1, \ldots, V_m\}$, called the *alphabet* of the propositional language. The domain of a propositional atom $V_i \in \Sigma$ is $\mathrm{dom}(V_i) = \{v_i, \overline{v}_i\}$, meaning that every atom may appear in positive (v_i) or negative ($\overline{v}_i$) form. A *literal* denotes a positive or negative form of a propositional atom. The set of formulas, that is, the language $\mathfrak{L}$ over Σ, can be defined recursively by use of the conjunction $\wedge$ (meaning the logical *and*), disjunction $\vee$ (the logical *or*)

Table 2 Comparison of the evaluations of the material conditional, the biconditional, and the conditional

$[\![\phi]\!]_\omega$	$[\![\psi]\!]_\omega$	$[\![\phi \Rightarrow \psi]\!]_\omega$	$[\![\phi \Leftrightarrow \psi]\!]_\omega$	$[\![(\psi\|\phi)]\!]_\omega$
true	*true*	*true*	*true*	*true*
true	*false*	*false*	*false*	*false*
false	*true*	*true*	*false*	*undefined*
false	*false*	*true*	*true*	*undefined*

The $[\![\cdot]\!]_\omega$ represents the evaluation of the respective formula given a world ω

and negation $\neg$ as usual. We use the symbols $\Rightarrow$ (material implication) and $\Leftrightarrow$ (equivalence), $\top$ (tautology) and $\bot$ (contradiction) with their usual semantics in this article. To improve readability, we omit the $\wedge$ operator and indicate negation by overlining. A *propositional interpretation* of Σ is a function I that assigns to each $V_i \in \Sigma$ a truth value, formally $I : \Sigma \rightarrow \{true, false\}$. To determine the truth value of a propositional formula ϕ given an interpretation I ($[\![\phi]\!]_I$), the recursive definition of the language is used. A *possible world* (Kripke 1972; Lewis 1986) is a propositional interpretation over $\mathfrak{L}$. It is syntactically represented by a complete conjunction of literals where every variable $V_i \in \Sigma$ appears exactly once only. The set of all *possible worlds* of Σ, that is, the set of worlds which are not ruled out by additional logical constraints, is called Ω. A *model* of a propositional formula $\phi \in \mathfrak{L}$ is a possible world ω that satisfies ϕ, written as $\omega \models \phi$. The set $Mod(\phi) = \{\omega | \omega \models \phi\}$ is the set of all worlds under which ϕ is evaluated to *true* (that is, $[\![\phi]\!]_\omega = true$). We overload $\models$ to act as relation between formulas $\phi, \psi \in \mathfrak{L}$ such that $\phi \models \psi$ if and only if $Mod(\phi) \subseteq Mod(\psi)$.

For two formulas $\phi, \psi \in \mathfrak{L}$, a *conditional* $(\psi|\phi)$ represents the defeasible rule "If ϕ, then normally ψ". A conditional is a trivalent logical entity that can be verified, falsified, or not applicable in a possible world $\omega \in \Omega$:

$$[\![(\psi|\phi)]\!]_\omega = \begin{cases} true & \text{iff} \quad \omega \models \phi\psi \qquad \text{(verification)} \\ false & \text{iff} \quad \omega \models \phi\overline{\psi} \qquad \text{(falsification)} \\ undefined & \text{iff} \quad \omega \models \overline{\phi} \qquad \text{(non} - \text{applicability)} \end{cases}$$

This trivalent evaluation of conditionals goes back to de Finetti (1974) and has been shown to be used in human reasoning (Wason 1968), instead of the evaluation of the material implication. Hence, this evaluation scheme is called a *defective* (Wason 1968) or *deFinetti truth table* (Baratgin et al. 2014). Note that this evaluation of conditionals deviates from the evaluation of the material conditional only in those cases where the premise is falsified, otherwise, it is identical to the evaluation of the material conditional, as is shown in Table 2.

Definition 1 *(Knowledge Base for Defeasible Reasoning)* A *knowledge base* $\mathcal{R} = (\mathcal{S}, \Delta)$ is a pair consisting of a set of propositional formulas $\mathcal{S} \subseteq \mathfrak{L}$ and a set of conditionals $\Delta \subseteq (\mathfrak{L}|\mathfrak{L})$. The propositional formulas encode strict knowledge, i.e., strict facts and formulas that are always true in the scenario described by the knowledge base.

The set of strict formulas $\mathcal{S}$ reduce the set of worlds to a set of *possible worlds* $\Omega(\mathcal{R}) \subseteq \Omega$ such that $\omega \in \Omega(\mathcal{R})$ if and only if $\omega \models \bigwedge_{\phi \in \mathcal{S}} \phi$. The conditional knowledge Δ describes defeasible connections between formulas which are normally true, i.e., their verification is more plausible than their falsification. To give appropriate semantics to conditionals, they are normally considered within richer structures, such as epistemic states in the sense of Halpern (2005), as are, for instance, ranking functions (Spohn 2012), possibility distributions (Dubois and Prade 2015), and probability distributions.

Besides strict knowledge, epistemic states also allow the representation of preferences, beliefs, and assumptions of an intelligent agent. Basically, an epistemic state allows one to compare formulas or worlds with respect to plausibility (or possibility, necessity and probability). This can be realized by preferential structures (see Sect. 3.4) like Ordinal Conditional Functions (Spohn 1988, 2012), which rank a world that falsifies a conditional $(\psi|\phi) \in \Delta$ to be less plausible than a world which, ceteris paribus, verifies or does not falsify this conditional. We illustrate these preliminaries with the following example that is inspired by the Suppression Task (Byrne 1989). This example will be used as a running example throughout the paper.

Example 1 We formalize the facts "she has an essay to write" (e), "she will not study late in the library" ($\bar{l}$) and the rule "if she has an essay to write, she will study late in the library" as strict and defeasible knowledge, respectively, which gives us the knowledge bases in Table 3. The sets of worlds that are possible for the different MP knowledge bases are given as follows.

$$\begin{aligned} \Omega(\mathcal{R}^{MP}_{CC}) &= \{el\} \\ \Omega(\mathcal{R}^{MP}_{CW}) &= \{el, e\bar{l}\} \\ \Omega(\mathcal{R}^{MP}_{FW}) &= \{el, \bar{e}l, \bar{e}\bar{l}\} \\ \Omega(\mathcal{R}^{MP}_{CFW}) &= \{el, e\bar{l}, \bar{e}l, \bar{e}\bar{l}\}. \end{aligned}$$

A set of conditionals $\Delta = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$ *tolerates* a conditional $(\psi|\phi)$ iff there is a possible world that verifies $(\psi|\phi)$ and does not falsify any conditional in Δ, i.e., there is an $\omega \in \Omega$ such that $\omega \models \phi\psi \wedge \bigwedge_{i=1}^{n}(\phi_i \Rightarrow \psi_i)$. A knowledge base $\mathcal{R} = (\mathcal{S}, \Delta)$ is *consistent* if and only if $\mathcal{S}$ is satisfiable and there is an ordered

Table 3 Knowledge bases ($\mathcal{R} = (\mathcal{S}, \Delta)$ with strict ($\mathcal{S}$) and defeasible knowledge (Δ) in accordance with Definition 1) that formalize the inference schemes given in Table 1 (see Example 1)

	Modus ponens	Modus tollens
Classical conditional	$\mathcal{R}^{MP}_{CC} = (\{e \Rightarrow l, e\}, \emptyset)$	$\mathcal{R}^{MT}_{CC} = (\{e \Rightarrow l, \bar{l}\}, \emptyset)$
Conditional weakening	$\mathcal{R}^{MP}_{CW} = (\{e\}, \{(l\|e)\})$	$\mathcal{R}^{MT}_{CW} = (\{\neg l\}, \{(l\|e)\})$
Fact weakening	$\mathcal{R}^{MP}_{FW} = (\{e \Rightarrow l\}, \{(e\|\top)\})$	$\mathcal{R}^{MT}_{FW} = (\{e \Rightarrow l\}, \{(\bar{l}\|\top)\})$
Conditional and fact weakening	$\mathcal{R}^{MP}_{CFW} = (\emptyset, \{\{(l\|e), (e\|\top)\}\})$	$\mathcal{R}^{MT}_{CFW} = (\emptyset, \{\{(l\|e), (\bar{l}\|\top)\}\})$

partition $(\Delta_0, \ldots, \Delta_m)$ of Δ such that each conditional in a partition i is tolerated by the union of all partitions with equal or higher order, i.e., of the set $\bigcup_{j=i}^{m} \Delta_j$.

3 Classes of Nonmonotonic Systems and Predictions

Instead of using the material implication, in commonsense reasoning, a conditional assertion commonly represents a notion of plausibility (Makinson 1994). The *classical consequence relation* $\models$ with $e \models l$ defines: if e is true, then l *must* be true, since this states that every model of e is a model of l according to the definition of $\models$ in Sect. 2. In contrast, the *nonmonotonic inference relation* (Kraus et al. 1990) uses $\mid\!\sim^{P}$, where $e \mid\!\sim^{P} l$ means, if e is true, then *typically* l is true as well (with $\mid\!\sim^{P}$ representing System P—see below). 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 which we will investigate in the following part.

For this comparative study or benchmark of approaches to nonmonotonic inference we selected the well-established, tested and tried approaches of System P (Sect. 3.1), Logic Programming (Sect. 3.2), Reiter Default Logic (Sect. 3.3) and, preferential reasoning with ranking functions (Sect. 3.4) which are generated by the inductive approaches of System Z or c-representations. Naturally, these approaches differ in a syntactical level as well as in their way of computing inferences. Using this broad set of approaches allows us to benchmark a wide area in the field of nonmonotonic inference rather than restricting ourselves to a small, confined area. We will present the formalisms and illustrate their workings by applying the MP case of the Suppression Task, translated into the respective syntax, as a running example in the respective sections. For reasons of space and possible loss of focus of this article, we refer the reader to the cited literature for a more thorough introduction to the approaches.

3.1 Class 1: System P

Adams' System P (Adams 1965) is probably one of the most famous formal systems for nonmonotonic reasoning characterizing it by six axioms listed in Table 4. A formula ψ follows preferentially in System P from a formula ϕ given a knowledge base $\mathcal{R} = (\mathcal{S}, \Delta)$ (written $\phi \mid\!\sim^{P}_{\mathcal{R}} \psi$) if and only if the extension of Δ with the conditional $(\overline{\psi}|\phi)$ is inconsistent (Dubois and Prade 1996; Goldszmidt and Pearl 1996). In the following section, we will show that in System P, no suppression effect will occur in general. Here, our knowledge base consists of conditional rules like "if she has an essay to write, she (normally) will study late in the library" $(l|e)$ and we write the minor premise "she has an essay to write" as a conditional $(e|\top)$. For our proof, we use the inference rules: 1. **Supraclassicality** (SCL) (Kraus et al. 1990), it states that the inference relation infers nonmonotonically from a premise what can be inferred in a classical way, 2. **Deduction** (DED) (Bochman 2001), it states that an inference relation allows to infer $\psi \Rightarrow \chi$ from a premise ϕ if its conclusion can be inferred from the conjunction of the original premise ϕ and the implication's

Table 4 Axioms of System P (see Adams 1965; Lehmann and Magidor 1992)

(REF)		For all	ϕ	it holds that	$\phi \mid\!\sim \phi$
(CM)	$\phi \mid\!\sim \psi$	and	$\phi \mid\!\sim \chi$	imply	$\phi\psi \mid\!\sim \chi$
(CUT)	$\phi \mid\!\sim \psi$	and	$\phi\psi \mid\!\sim \chi$	imply	$\phi \mid\!\sim \chi$
(RW)	$\phi \mid\!\sim \psi$	and	$\psi \models \chi$	imply	$\phi \mid\!\sim \chi$
(LLE)	$\phi \equiv \psi$	and	$\psi \mid\!\sim \chi$	imply	$\phi \mid\!\sim \chi$
(OR)	$\phi \mid\!\sim \chi$	and	$\psi \mid\!\sim \chi$	imply	$(\phi \vee \psi) \mid\!\sim \chi$

Reflexivity (REF) states that everything that is already known is inferable; Cautious Monotony (CM) states that inferred knowledge can be added to the premises without rendering that conclusion impossible; Cut (CUT) states that the inferences of a premise include those inferences of this premise together with valid inferences; Right Weakening (RW) states that a nonmonotonic inference includes the classical inferences from the conclusion; Left Logical Equivalence (LLE) states that the inferences from two semantically equivalent premises are identical; Or (OR) states that if a conclusion can be drawn from two premises independently, it also can be drawn from the disjunction of this premises

premise ψ, and 3. **Modus Ponens in the consequence** (MPC) (Kraus et al. 1990), it states that if both a material implication and its premise can be inferred from a formula, the conclusion of the material implication can be inferred directly.

$$\begin{array}{llllll} \text{(SCL)} & & \phi \models \psi & & \text{implies} & \phi \mid\!\sim \psi \\ \text{(DED)} & & \phi \wedge \psi \mid\!\sim \chi & & \text{implies} & \phi \mid\!\sim (\psi \Rightarrow \chi) \\ \text{(MPC)} & \phi \mid\!\sim \psi & \text{and} & \phi \mid\!\sim (\psi \Rightarrow \chi) & \text{imply} & \phi \mid\!\sim \chi \end{array}$$

All these inference rules are derivable from the axioms of System P; Fig. 1 illustrates the inference chains from the axioms to the respective properties. If a conditional $(\psi|\phi)$ is an element of the knowledge base, then the knowledge base will become inconsistent by adding the conditional $(\overline{\psi}|\phi)$. Hence, for each $(\psi|\phi) \in \Delta$, we have $\phi \mid\!\sim^{P}_{\mathcal{R}} \psi$, a property known as *Direct Inference* (DI) (Łukasiewicz 2005). So, for each $(\psi|\phi) \in \Delta$, (DI) gives us $\phi \mid\!\sim^{P}_{\mathcal{R}} \psi$, and from this, we obtain $\top \mid\!\sim^{P}_{\mathcal{R}} (\phi \Rightarrow \psi)$ by applying (DED). If we also have $(\phi|\top) \in \Delta$, then through (DI) and (MPC) we obtain

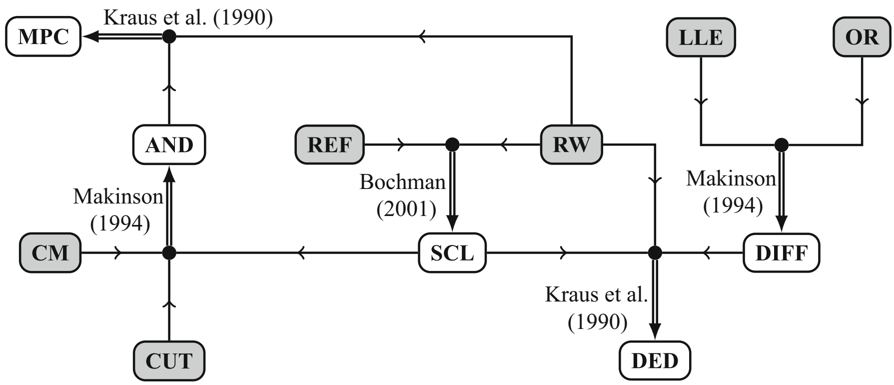

Fig. 1 Important inference properties (*white*) derivable from the axioms (*gray*) of System P, where ⇒ indicates an implication from the ingoing properties to the outgoing property, annotated with the reference for the proof

$$\dfrac{\dfrac{(\phi|\top)\in\Delta}{\top\mid\!\sim\phi}\,\text{(DI)} \qquad \dfrac{\dfrac{\dfrac{(\psi|\phi)\in\Delta}{\phi\mid\!\sim\psi}\,\text{(DI)}}{\top\wedge\phi\mid\!\sim\psi}\,\text{(LLE)}}{\top\mid\!\sim\phi\Rightarrow\psi}\,\text{(DED)}}{\top\mid\!\sim\psi}\,\text{(MPC)}$$

(a) Defeasible knowledge about ϕ

$$\dfrac{\dfrac{\top\models\phi}{\top\mid\!\sim\phi}\,\text{(SCL)} \qquad \dfrac{\dfrac{\dfrac{(\psi|\phi)\in\Delta}{e\mid\!\sim\psi}\,\text{(DI)}}{\top\wedge\phi\mid\!\sim\psi}\,\text{(LLE)}}{\top\mid\!\sim\phi\Rightarrow\psi}\,\text{(DED)}}{\top\mid\!\sim\psi}\,\text{(MPC)}$$

(b) Strict knowledge about ϕ

Fig. 2 Inference of ψ from having $(\psi|\phi)$ in the knowledge base using System P rules given defeasible (*left*) or strict (*right*) knowledge about ϕ. **a** Defeasible knowledge about ϕ. **b** Strict knowledge about ϕ

$\top\mid\!\sim^{P}_{\mathcal{R}}\psi$, no matter which other knowledge is present in the knowledge base. If we have strict knowledge $\phi\in\mathcal{S}$, then with (SCL) and (MPC) we also obtain $\top\mid\!\sim^{P}_{\mathcal{R}}\phi$. Hence, System P does not demonstrate the suppression effect in general. Figure 2 graphically illustrates the steps of the proofs.

Example 2 *(System P entailment)* We illustrate this with the standard P entailment as defined above. We use the knowledge base $\mathcal{R}^{MP}_{P}=(\emptyset,\{(l|e),(l|o),(e|\top)\})$ as an example. $\Delta^{MP}_{P}\cup\{(l|\top)\}$ is consistent as, e.g., the world elo verifies all conditionals. On the other hand, $\Delta^{MP}_{P}\cup\{(\bar{l}|\top)\}$ is inconsistent since there is no world that verifies at least one of the conditionals and accepts all the others. Therefore, we have $\top\mid\!\sim^{P}_{\mathcal{R}^{MP}_{P}}l$ as well as $\top\mid\!\not\sim^{P}_{\mathcal{R}^{MP}_{P}}\bar{l}$, and hence "she will study late in the library". And thus we do not find any suppression effect on the inferences even if information about the necessary precondition of the library being open is present.

3.2 Class 2: Logic Programming

Logic programming is a inference system where logical rules are used to represent (strict) rules and the knowledge base is composed of such rules. In its most basic form, a rule is a Horn clause, that is, the conclusion (often called the "head"in this context) of the rule is a positive literal v with $V\in\Sigma$ and the premise ("body") of the rule is a conjunction of positive literals $w_1\wedge\cdots\wedge w_m$ with $w_i\in\Sigma$ for $1\leq i\leq m$. A relaxation of this system is to allow negative literals in the premise.

This language is used by Stenning and Lambalgen (2008) and Dietz et al. (2012) to model the Suppression Task. The basic idea is that a conditional in natural language forms a specific structure—a "license for inference", i.e., rules like $v\leftarrow w_1\wedge\cdots\wedge w_m\wedge\neg ab_i$, where ab_i is an abnormality for the rule which, if positive, inhibits the conclusion of v. Such programs are evaluated under weak completion semantics (Hölldobler and Kencana Ramli 2009b). The following replacement steps are executed on the program:

1. Replace each set of clauses with identical head v by a single clause with head v and a disjunction of all body literals of the clauses. Example: $(v\leftarrow w_1),\ldots,(v\leftarrow w_n)$ is replaced by $v\leftarrow w_1\vee\ldots\vee w_n$.
2. Replace each occurrence of $\leftarrow$ by $\Leftrightarrow$.

The resulting set of equivalences is the so called *weak completion* of the program.[1] 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 2009a), that means, the intersection of two models is also a model. This property entails the existence of least models. In the following, we outline the idea using three-valued Łukasiewicz Semantics (Łukasiewicz 1920). Truth values are $\top$ (*true*), $\bot$ (*false*) and U (*unknown*). A *three-valued interpretation* I maps a formula F to $\{\top, \bot, \mathsf{U}\}$ (Dietz et al. 2012). Each formula is evaluated according to the truth tables of Łukasiewicz (Dietz et al. 2012). An interpretation is written as a pair $I = \langle I^\top, I^\bot \rangle$ of disjoint sets of atoms where $I^\top$ is the set of all atoms that are mapped to $\top$ by I; while $I^\bot$ is the set of all atoms that are mapped to $\bot$ by I. Atoms that are not an element of $I^\top \cup I^\bot$ are mapped to U. We write $I(F) = \top$ for a formula F that is evaluated to be true under the interpretation I. We call M a *model* of a logic program $\mathcal{P}$ if it is an interpretation evaluating each clause occurring in $\mathcal{P}$ to be $\top$. Weak completion semantics considers weakly completed logic programs and reasons with respect to the least Ł-models of these programs. The least Ł-model of a weakly completed logic program $\widehat{\mathcal{P}}$ can be obtained by computing the least fixed point of the following semantic operator, as explained in Stenning and van Lambalgen (2008):

$$\Phi_P(\langle I^\top, I^\bot \rangle) = \langle J^\top, J^\bot \rangle$$

with

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

The model relation $\models_{wc}$ for a program $\mathcal{P}$ is defined by $\mathcal{P}$: $\mathcal{P} \models_{wc} F$ iff the formula F holds in the least Ł-model of the weakly completed program $\widehat{\mathcal{P}}$. Another formulation is that for all formulas F we have $\mathcal{P} \models_{wc} F$ if and only if for the least model lm of the weakly completed $lm_{wc}\mathcal{P}(F) = \top$ is evaluated as true.

Example 3 We illustrate logic programming with the Suppression Task in its MP and MT cases here. Suppose we have rules "If she has an essay to write *and nothing* $abnormal_1$ *happens*, she will study late in the library" and "If the library is open *and nothing* $abnormal_2$ *happens*, she will study late in the library". The possible abnormal situations for these cases may differ, leading to the difference between the two cases. Furthermore, "the library not being open" is an $abnormal_1$ situation (since she cannot be in the library if it is closed) and "not having an essay to write" is an $abnormal_2$ situation (since then she has no reason to stay in the library). An additional premise, either "she has an essay to write" or "she does not study late in the library", is added to this knowledge with respect to the task (MP or MT

[1] The expression weak completion is used to denote the difference to completion processes that consider a mapping of undefined atoms to $\bot$ (cp. Dietz et al. 2015).

respectively) at hand. The complete implementation of the Suppression Task for modus ponens and modus tollens in this language is:

$$\mathcal{P}_{MP} = \left\{ \begin{array}{ccc} l & \leftarrow & e \wedge \overline{ab}_1, \\ l & \leftarrow & o \wedge \overline{ab}_2, \\ ab_1 & \leftarrow & \overline{o}, \\ ab_2 & \leftarrow & \overline{e}, \\ e & \leftarrow & \top \end{array} \right\}, \mathcal{P}_{MT} = \left\{ \begin{array}{ccc} l & \leftarrow & e \wedge \overline{ab}_1, \\ l & \leftarrow & o \wedge \overline{ab}_2, \\ ab_1 & \leftarrow & \overline{o}, \\ ab_2 & \leftarrow & \overline{e}, \\ l & \leftarrow & \bot \end{array} \right\} \quad (1)$$

We apply weak completion to the two programs and obtain the equivalences

$$\begin{array}{l} \widehat{\mathcal{P}}_{MP} = \left\{ \begin{array}{ccc} l & \Leftrightarrow & (e \wedge \overline{ab}_1) \vee (o \wedge \overline{ab}_2), \\ ab_1 & \Leftrightarrow & \overline{o}, \\ ab_2 & \Leftrightarrow & \overline{e}, \\ e & \Leftrightarrow & \top \end{array} \right\}, \\ \widehat{\mathcal{P}}_{MT} = \left\{ \begin{array}{ccc} l & \Leftrightarrow & (e \wedge \overline{ab}_1) \vee (o \wedge \overline{ab}_2), \\ ab_1 & \Leftrightarrow & \overline{o}, \\ ab_2 & \Leftrightarrow & \overline{e}, \\ l & \Leftrightarrow & \bot \end{array} \right\} \end{array} \quad (2)$$

Here, the least model for $\mathcal{P}_{MP}$ is $(\{e\}, \{ab_2\})$ and the least model for $\mathcal{P}_{MT}$ is $(\emptyset, \{l\})$. So we have neither $l \in I^\top$ nor $l \in I^\perp$ in the MP case and hence no information about "being in the library" is present. To conclude, we obtain that neither the MP nor the MT inferences are drawn when implementing the Suppression Task as logic programs.

3.3 Class 3: Reiter Default Logic

Reiter default logic (Reiter 1980) realizes the idea that often rules which we know about the world can only be "almost always" true with some exceptions. For instance, "Most birds fly" is a valid rule, though there are exceptions like the penguins. Given a particular bird, we will conclude that it flies as long as we do not have the additional information that it does not. In general, rules in Reiter default logic can be applied "in the absence of any information to the contrary" (Reiter 1980). This leads to the typical structure of a default rule $\delta = \frac{pre(\delta):just(\delta)}{cons(\delta)}$. A default δ is comprised of a precondition (the formula $pre(\delta)$), a set of justifications $just(\delta)$ and a set of consequences $cons(\delta)$. A set of default rules $\mathcal{D} = \{\delta_1, \delta_2, \ldots, \delta_n\}$ together with a classical logical background theory $\mathcal{W}$ form a Reiter default theory $(\mathcal{W}, \mathcal{D})$.

The consequences of a default can be inferred if the precondition is satisfied and the justifications can be assumed to be consistent. Formally, a default δ is *applicable* to a deductively closed set $Cn(\mathcal{A})$ if and only if $pre(\delta) \in Cn(\mathcal{A})$ and $\neg B \notin Cn(\mathcal{A})$ for every $B \in just(\delta)$.

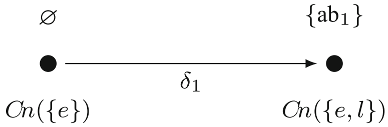

Fig. 3 The process tree visualizing the Reiter default processes for Example 4 where each vertex is annotated with the respective sets *In* (*below*) and *Out* (*above*)

A finite sequence of defaults $(\delta_{\Pi_1}, \ldots, \delta_{\Pi_m})$, $\delta_{\Pi_i} \in D$ for all $1 \leq i \leq m$ is called a *(default) process* Π (Antoniou 1997) with the two sets $In(\Pi) = Cn(W \cup \{cons(\delta) | \delta \in \Pi\})$ and $Out(\Pi) = \{\neg A | A \in just(\delta), \delta \in \Pi\}$ if and only if each default δ is applicable to the *In*-set of the previous defaults. A process is *successful* if and only if $In(\Pi) \cap Out(\Pi) = \emptyset$ and *closed* if and only if every $\delta \in D$ that is applicable to $In(\Pi)$ is an element of Π. The set E is called an *extension* if and only if there exists a successful and closed process Π with $E = In(\Pi)$. A formula ψ follows with Reiter default logic from a set of formulas $\mathcal{W}$ given a set of defaults $\mathcal{D}$ (written $\mathcal{W} {\mid\!\sim}^{Reiter}_{\mathcal{D}} \psi$), if and only if $\psi \in \bigcap \mathcal{E}$ with $\mathcal{E}$ the set of all extensions E of the Reiter default theory $(\mathcal{W}, \mathcal{D})$.

Example 4 *(Reiter)* Consider the Reiter default theory $(\mathcal{W}^{MP}, \mathcal{D}_P)$ with $\mathcal{D}_P = \left\{\delta_1 : \frac{e:\neg ab_1}{l}, \delta_2 : \frac{o:\neg ab_2}{l}, \delta_3 : \frac{\neg o:ab_1}{ab_1}, \delta_4 : \frac{\neg e:ab_2}{ab_2}\right\}$ and $\mathcal{W}^{MP} = \{e\}$, that is obtained from transforming the rules of the logic program for modus ponens in (1) into Reiter defaults. Figure 3 shows the process tree for this theory. It can be seen that there is only one extension, $Cn(\{e, l\})$, and since $l \in Cn(\{e, l\})$, we can infer that "she will study late in the library", formally $e {\mid\!\sim}^{Reiter}_{\mathcal{D}_P} l$.

3.4 Class 4: Ranking Models

Reasoning with Ordinal Conditional Functions (OCF, ranking functions Spohn 1988, 2012) is preferential reasoning (see Makinson 1994) with a preference relation induced by the plausibility ranking from an OCF κ. An OCF κ is a function which assigns to each world $\omega \in \Omega$ a degree of implausibility $\kappa : \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$ such that there are maximally plausible worlds, that is, worlds with a rank of 0, and thus $\kappa^{-1}(0) \neq \emptyset$. This ranking function then induces the preference relation of a preferential model $(\Omega, \models, <_\kappa)$ (Makinson 1994) so that $\omega <_\kappa \omega'$ if and only if $\kappa(\omega) < \kappa(\omega')$. We define the rank of a formula $\phi \in \mathfrak{L}$ to be the rank of the most plausible world that satisfies ϕ, formally $\kappa(\phi) = \min\{\kappa(\omega) \mid \omega \models \phi\}$. and the rank of a conditional to be the rank of the verification of the conditional normalized by the rank of the premise, formally $\kappa(\psi|\phi) = \kappa(\phi\psi) - \kappa(\phi)$.

This directly gives us that inferring preferentially with a preference relation induced by an OCF is an open world inference: Let $\Sigma' = \{U_1, \ldots, U_k\} \subseteq \Sigma$ be a set of variables and let $\mathbf{o}$ and $\mathbf{o}'$ be conjunctions of instantiations of variables of Σ'.

Then, both of them can be compared with respect to their plausibility using OCF by marginalization, i.e., by comparing the ranks of the respective formulas. We have $\kappa(\mathbf{o}) = \min\{\kappa(\omega) | \omega \models \mathbf{o}\}$ (likewise for $\mathbf{o}'$). By explicitly setting non-instantiated (or unknown) variables to *false*, OCF can be extended to accept the closed world assumption, if necessary.

With their plausibility ranking, OCFs give semantics to conditionals such that an OCF κ satisfies a conditional if and only if the verification of the conditional is strictly more plausible than its falsification, formally

$$\kappa \models (\psi|\phi) \quad \text{if and only if} \quad \kappa(\phi\psi) < \kappa(\phi\overline{\psi}).$$

We define the inference relation induced by an OCF in accordance with the preferential model $(\Omega, \models, <_\kappa)$. In other words, ψ can be κ-inferred from ϕ if and only if κ satisfies the conditional $(\psi|\phi)$ (see Kern-Isberner and Eichhorn 2012 for a formal proof).

$$\phi \mathrel{|\!\!\sim}^{\kappa} \psi \quad \text{if and only if} \quad \kappa(\phi\psi) < \kappa(\phi\overline{\psi}) \quad \text{if and only if} \quad \kappa \models (\psi|\phi) \tag{3}$$

This *ranking inference* is a prototypical instantiation of preferential reasoning in the scope of System P. Additionally, skeptical inference over *all* $\mathcal{R}$-admissible OCFs is identical with System P inference (Goldszmidt and Pearl 1996).

An OCF κ is admissible with respect to a knowledge base $\mathcal{R} = (\mathcal{S}, \Delta)$ if and only if it satisfies all conditionals in the knowledge base, that is, $\kappa \models (\psi|\phi)$ for all $(\psi|\phi) \in \Delta$. We write $\kappa_{\mathcal{R}}$ to indicate that the OCF is admissible with respect to $\mathcal{R}$. The inference relation $\mathrel{|\!\!\sim}^{\kappa_{\mathcal{R}}}$ of an OCF admissible to a knowledge base $\mathcal{R}$ is defined according to (3); this inference relation describes an inference based on the background knowledge $\mathcal{R}$.

In the following part, we will present two approaches to generate OCFs that are admissible with respect to a given knowledge base: Pearl's System Z (1990) and c-representations (Kern-Isberner 2001, 2004).

System Z generates the unique Pareto-minimal OCF that is admissible with respect to the knowledge base (that is, every other admissible OCF assigns a larger rank to at least one world ω), by iteratively applying the following notion of tolerance between conditionals and sets of conditionals to the knowledge base. A set of conditionals $\Delta = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$ *tolerates* a conditional $(\psi|\phi)$ if and only if there is a world $\omega \in \Omega$ that verifies $(\psi|\phi)$ and does not falsify any conditional in Δ, formally

$$\exists\, \omega : \omega \models \left(\phi\psi \wedge \bigwedge_{i=1}^{n} \phi_i \Rightarrow \psi_i\right).$$

Hereby, the conditional knowledge base is partitioned into (with respect to set inclusion) maximal sets of mutual tolerance $(\Delta_0, \Delta_1, \ldots \Delta_k)$ according to Algorithm 1. We define an auxiliary function $Z : (\mathfrak{L}|\mathfrak{L}) \to \mathbb{N}$ as the function which assigns to each conditional $(\psi|\phi) \in \Delta$ the index of the partition it is an element of.

Algorithm 1 An algorithm to evaluate the consistency for a given set of conditionals $\Delta = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$. It is used to construct the partitions needed for System Z [45].

```
INPUT  : Δ = {(ψ1|φ1), ..., (ψn|φn)}
OUTPUT : Ordered partition (Δ0, Δ1, ...Δk) if Δ is consistent,
         NULL otherwise

BEGIN
  INT i:=0;
  WHILE(Δ ≠ ∅) DO
    Δi := {(B|A) | (B|A) ∈ Δ and Δ tolerates (B|A)};
    IF (Δi ≠ ∅)
    THEN
      Δ:=Δ \ Δi;
      i:=i+1;
    ELSE
      RETURN NULL;  // Δ is inconsistent
    ENDIF
  END
  RETURN Δ = ⟨Δ0, Δ1, ..., Δi⟩;
END
```

Formally, $Z(\psi|\phi) = j$ if and only if $(\psi|\phi) \in \Delta_j$ and Z being undefined for conditionals which are not in the knowledge base. With this function, the ranking function $\kappa^Z_{\mathcal{R}}$ is defined such that the rank of a world is 0 if and only if the world does not falsify any conditional. If the world does falsify conditionals, the rank of the world is set with respect to the maximal index of the falsified conditionals, formally

$$\kappa^Z_{\mathcal{R}}(\omega) = \begin{cases} 0 & \text{if} \quad \omega \models \bigwedge_{\mathrm{i=1}}^{\mathrm{n}} \phi_{\mathrm{i}}\psi_{\mathrm{i}} \\ \max\{Z(\psi_i|\phi_i)|\omega \models \phi_i\overline{\psi_i}\} + 1 & \text{otherwise} \end{cases} \tag{4}$$

Example 5 (System Z) We illustrate System Z with the knowledge base $\mathcal{R}^{MP}_P = (\emptyset, \Delta^{MP}_P = \{(l|e), (l|o), (e|\top)\})$. During the consistency test, all conditionals from the knowledge base are put into $\Delta^{MP}_{P,0}$ because the world *elo* verifies every conditional in the knowledge base, hence every conditional is tolerated by Δ^{MP}_P. From the resulting ranking function, we find $\kappa^Z_{\mathcal{R}^{MP}_P}(l) = 0 < 1 = \kappa^Z_{\mathcal{R}^{MP}_P}(\bar{l})$, so l is strictly more plausible than $\bar{l}$ and thus System Z infers that “she will study late in the library”, formally $e \mid\!\sim^{\kappa^Z_{\mathcal{R}}} l$.

c-representations Other than calculating a single OCF-model of the knowledge base, the approach of c-representations (Kern-Isberner 2001, 2004) provides a schema for $\mathcal{R}$-admissible OCFs. Here, each conditional $(\psi_i|\phi_i) \in \Delta = \{(\psi_1|\phi_1), \ldots, (\psi_n|\phi_n)\}$ is associated with an integer impact κ_i^-. The rank of a world ω is determined by summing up the impacts of conditionals in Δ which are falsified by ω, formally

$$\kappa^c_{\mathcal{R}}(\omega) = \sum_{i=1}^{n} \begin{cases} \kappa_i^- & \text{iff } \omega \models \phi_{\mathrm{i}}\overline{\psi_{\mathrm{i}}} \\ 0 & \text{otherwise} \end{cases} = \sum_{\omega \models \phi_i\overline{\psi_i}} \kappa_i^-, \tag{5}$$

where the impacts are chosen such that $\kappa_{\mathcal{R}}^c \models \Delta$, which is the case if and only if

$$\kappa_i^- > \min_{\omega \models \phi_i \psi_i} \left\{ \sum_{\substack{\omega \models \phi_j \overline{\psi_j} \\ j \neq i}} \kappa_j^- \right\} - \min_{\omega \models \phi_i \overline{\psi_i}} \left\{ \sum_{\substack{\omega \models \phi_j \overline{\psi_j} \\ j \neq i}} \kappa_j^- \right\} \quad \forall 1 \leq i \leq n \qquad (6)$$

Note that since (6) is a system of inequalities, there are infinite solutions which satisfy (6) and each solution $\kappa^- = (\kappa_1^-, \ldots, \kappa_n^-)$ generates a ranking function according to (5). Mimicking System Z, we here concentrate on solutions that generate Pareto-minimal OCFs $\kappa_{\mathcal{R}}^c$ with respect to (5). Thorn et al. (2015) discusses how to generate a respective unique solution κ^- if this is necessary, whilst (Beierle et al. 2016) discusses how a skeptical inference relation over all solutions of (6) can be realized.

Example 6 (c-representations) To illustrate c-representations, we again use the knowledge base $\mathcal{R}_P^{MP} = (\emptyset, \Delta_P^{MP} = \{(l|e), (l|o), (e|\top)\})$. The system of inequalities resulting from $\mathcal{R}_P^{MP}$ can be solved minimally with the values $\kappa_1^- = 1, \kappa_2^- = 0$ and $\kappa_3^- = 1$. For the OCF resulting from this solution, we find $\kappa_{\mathcal{R}_P^{MP}}^c(l) = 0 < 1 = \kappa_{\mathcal{R}_P^{MP}}^c(\bar{l})$. Therefore, "studying late in the library" is strictly more plausible than "not studying late in the library" and hence inference with c-representations gives us l, or $e \mid\!\sim^{\kappa_{\mathcal{R}}^c} l$.

3.5 Modeling Background Knowledge with Conditionals

It has been shown in Ragni et al. (2016) that, when using the plain modeling $\mathcal{R}_P$, neither System P nor the ranking approaches show the suppression effect, in contrast to logic programming with weak completion semantics (Stenning and Lambalgen 2008). The programs used in Stenning and Lambalgen (2008) and Dietz et al. (2012) encode the two cases differently. As argued in Ragni et al. (2016), for the modus ponens, it is possible to mimic this effect of weak completion semantics with conditional knowledge bases. We made background knowledge an explicit part of the knowledge base and modeling the MP and MT case for the other systems.

Hence, the suppression effect can also occur—not as a consequence of a suitable reasoning mechanism but rather due to different background knowledge that is triggered by the additional information included in the knowledge base for transparency.

To analyze this effect of implicit knowledge, we will consider the different variables as being connected in the background knowledge. So apart from the plain approach used to illustrate the systems, we will also consider that "having an essay to write" and "the library being open" are either connected in the premise ($\mathcal{R}_{CP}$), as a necessary condition for "studying late in the library" ($\mathcal{R}_{NC}$), or both of these ($\mathcal{R}_{WC}$). Table 5 in Sect. 4 gives the formal realizations of these knowledge bases.

Table 5 Realizations of the Suppression Tasks as knowledge bases for the System P and ranking approaches

	Knowledge base
Plain approach	$\mathcal{R}_P = (\emptyset, \{\lambda_1 : (l\|e), \lambda_2 : (l\|o)\})$
Connecting premise	$\mathcal{R}_{CP} = (\emptyset, \{\lambda_5 : (l\|eo)\})$
Necessary condition	$\mathcal{R}_{NC} = (\emptyset, \{\lambda_6 : (eo\|l)\})$
Weak completion	$\mathcal{R}_{WC} = (\emptyset, \{\lambda_5 : (l\|eo), \lambda_6 : (eo\|l)\})$

For the MP case, we extend the default knowledge base of $\mathcal{R}_\bullet$ with the conditional $\lambda_3 : (e|\top)$ and, for the MT case, we extend the default knowledge base of $\mathcal{R}_\bullet$ with the conditional $\lambda_4 : (\bar{l}|\top)$ for all $\bullet \in \{P, CP, NC, WC\}$

4 Predictions of Formal Systems on the Suppression Task

After introducing different classes of established nonmonotonic systems, in this section we examine the possible inferences these systems draw for the Suppression Task. A previous analysis (Ragni et al. 2016) showed that for a plain (or "literal") realization ($\mathcal{R}_P$), neither System P nor the ranking approaches show the desired suppression effect of the MP inferences, in contrast to logic programming with weak completion semantics (Stenning and Lambalgen 2008; Dietz et al. 2012).

This holds due to a different encoding of the programs (1) (see Dietz et al. 2012) by using background knowledge. In the following, we show that by using knowledge bases that include an encoding of additional background knowledge, a suppression effect is possible within these approaches. Hence, there are two possible explanation patterns: One is based on the inference mechanism and the other on the appropriate knowledge base.

Most of the approaches described in Sect. 3 infer knowledge from a conditional knowledge base. The knowledge bases presented in Table 5 are based on and extend the modeling proposed in Ragni et al. (2016) especially to the extent that here we do not only work with the MP, but also the MT case of the Suppression Task. Here, the *plain* realization $\mathcal{R}_P$ models the conditional statements "if she has an essay to write, she will normally study late in the library" and "if the library is open, she will normally study late in the library" as two conditionals as it has been done for System P and the ranking models in Sect. 3.4. In the *connecting premise* approach $\mathcal{R}_{CP}$, the two premises of the statements are joined into a single premise since the second information "if the library is open, she will study late in the library" leads to considering o as an additional prerequisite for studying in the library, so this approach uses the conditional "if she has an essay to write and the library is open, she will normally study late in the library". Note that this corresponds roughly to the way knowledge is encoded in $\mathcal{P}_{MP}$ in (1). Another alternative to model this is to treat both premises as necessary conditions for being in the library, hence the *necessary condition* modeling uses the conditional "if she is in the library, she has an essay to write and the library is open". Finally, going for the equivalence operation in the weak completion semantics, we model that the premises are necessary and sufficient for being in the library by using both conditionals from the two previous realizations for the *weak completion* modeling. All these knowledge bases can be used for the MP case of the Suppression Task by adding the conditional $(e|\top)$ to the conditional knowledge base (we indicate this by the superscript *MP*), or, by adding the conditional $(\bar{l}|\top)$, for the MT case of the Suppression Task (indicated by the superscript *MT*):

$$\mathcal{R}_{\bullet}^{MP} = \mathcal{R}_{\bullet} \cup \{(e|\top)\}$$
$$\mathcal{R}_{\bullet}^{MT} = \mathcal{R}_{\bullet} \cup \{(\bar{l}|\top)\}$$

Class 1: System P

As already discussed, from $\mathcal{R}_P^{MP}$ System P will draw the MP inference by applying the properties **Deduction** and **Modus ponens in the consequence**. For the same reason, this system licenses for the MT inference from $\mathcal{R}_P^{MT}$. For all other realizations, the knowledge bases $\mathcal{R}_{\bullet}^{MP} \cup \{(\bar{l}|\top)\}$ respectively $\mathcal{R}_{\bullet}^{MT} \cup \{(e|\top)\}$ are consistent, which implies that l and $\bar{e}$ are not inferred. So adding the additional information about the library will suppress System P inferences for modus ponens and modus tollens only if this new information (that is, the library being open) is somehow connected to the information already present (her being in the library if she has an essay to write).

Class 3: Reiter Default Logic

For Reiter Default Logic we used the established realization of the Suppression Task from logic programming as *plain* approach $\mathcal{D}_P$. Furthermore we translated the other three knowledge bases (CP, NC and WC) from the System P approach into normal defaults, which resulted in the default sets $\mathcal{D}_{CP}, \mathcal{D}_{NC}$, and $\mathcal{D}_{WC}$ given in Table 6. This default knowledge is combined with the knowledge $\mathcal{W}^{MP} = \{e\}$ for the MP case and $\mathcal{W}^{MT} = \{\bar{l}\}$. We explicated the inferences for the default theory $(\mathcal{W}^{MP}, \mathcal{D}_P)$ in Example 4. In the MT case no default is applicable, because $\bar{l} \notin pre(\delta)$ for all $\delta_i \in \mathcal{D}_P$. Therefore, the only extension is $Cn(\mathcal{W}^{MT}) = Cn(\{\bar{l}\})$. For all other default sets there is no default applicable neither in the MP case nor in the MT case. In the MP case, we have $\{e\}$ but need $e \wedge o$ to apply the default δ_5 and in the MT case, we have $\{\bar{l}\}$ but δ_6 needs l. So, we have the extensions $Cn(\mathcal{W}^{MP}) = Cn(\{e\})$ and $Cn(\mathcal{W}^{MT}) = Cn(\{\bar{l}\})$ for the knowledge bases CP, NC, and WC.

Table 6 Realizations of the Suppression Task in Reiter default logic

	Reiter default logic
Plain approach	$\mathcal{D}_P = \left\{\delta_1 : \frac{e : \neg ab_1}{l}, \delta_2 : \frac{o : \neg ab_2}{l}, \delta_3 : \frac{\neg o : ab_1}{ab_1}, \delta_4 : \frac{\neg e : ab_2}{ab_2}\right\}$
Connecting premise	$\mathcal{D}_{CP} = \left\{\delta_5 : \frac{eo : l}{l}\right\}$
Necessary condition	$\mathcal{D}_{NC} = \left\{\delta_6 : \frac{l : eo}{eo}\right\}$
Weak completion	$\mathcal{D}_{WC} = \left\{\delta_5 : \frac{eo : l}{l}, \delta_6 : \frac{l : eo}{eo}\right\}$

For the MP case, we assume $\mathcal{W}^{MP} = \{e\}$ and for the MT case, we assume $\mathcal{W}^{MT} = \{\bar{l}\}$

Table 7 Ranking functions generated with System Z and c-representations for $\mathcal{R}_P^{MP}$ in Examples 5 and 6

ω	elo	$el\overline{o}$	$e\overline{l}o$	$e\overline{l}\overline{o}$	$\overline{e}lo$	$\overline{e}l\overline{o}$	$\overline{e}\overline{l}o$	$\overline{e}\overline{l}\overline{o}$
$\kappa^Z_{\mathcal{R}_P^{MP}}(\omega)$	0	0	1	1	1	1	1	1
$\kappa^c_{\mathcal{R}_P^{MP}}(\omega)$	0	0	1	1	1	1	1	1

Class 4: Ranking Models

For the four knowledge bases, System Z and c-representations show the same behavior as System P both in the MP and MT case, so we obtain that the strong connection of the additional information suppresses the inference for these strictly stronger systems as well: Examples 5 and 6, with the complete ranking functions given in Table 7, give the explicit calculations for the knowledge bases $\mathcal{R}_P^{MP}$ and $\mathcal{R}_P^{MT}$ for these approaches. For CP, NC, and WC the ranking functions inductively generated by System Z and c-representations are listed in Table 8. From this table we obtain that we have $\kappa^{\circ}_{c\mathcal{R}_\bullet^{MP}}(l) = 0 = \kappa^{\circ}_{c\mathcal{R}_\bullet^{MP}}(\overline{l})$ in the MP case and $\kappa^{\circ}_{c\mathcal{R}_\bullet^{MP}}(e) = 0 = \kappa^{\circ}_{c\mathcal{R}_\bullet^{MP}}(\overline{e})$ for the MT case for all $\circ \in \{c, Z\}$ and all $\bullet \in \{CP, NC, WC\}$ (cp. Table 5), and thus we have a suppression effect.

Overview of each system's prediction

Table 9 summarizes the inferences that can be drawn from the different knowledge bases using the presented approaches.

Table 8 Ranking functions generated with System Z and c-representations for the different knowledge bases with different forms of strict or defeasible knowledge CP, NC, and WC (cp. Table 5)

ω	elo	$el\overline{o}$	$e\overline{l}o$	$e\overline{l}\overline{o}$	$\overline{e}lo$	$\overline{e}l\overline{o}$	$\overline{e}\overline{l}o$	$\overline{e}\overline{l}\overline{o}$
$\kappa^Z_{\mathcal{R}_{CP}^{MP}}(\omega)$	0	0	1	0	1	1	1	1
$\kappa^c_{\mathcal{R}_{CP}^{MP}}(\omega)$	0	0	1	0	1	1	1	1
$\kappa^Z_{\mathcal{R}_{CP}^{MT}}(\omega)$	1	1	2	0	1	1	0	0
$\kappa^c_{\mathcal{R}_{CP}^{MT}}(\omega)$	1	1	2	0	1	1	0	0
$\kappa^Z_{\mathcal{R}_{NC}^{MP}}(\omega)$	0	1	0	0	1	1	1	1
$\kappa^c_{\mathcal{R}_{NC}^{MP}}(\omega)$	0	1	0	0	2	2	1	1
$\kappa^Z_{\mathcal{R}_{NC}^{MT}}(\omega)$	1	1	2	0	1	1	0	0
$\kappa^c_{\mathcal{R}_{NC}^{MT}}(\omega)$	1	1	2	0	1	1	0	0
$\kappa^Z_{\mathcal{R}_{WC}^{MP}}(\omega)$	0	1	1	0	1	1	1	1
$\kappa^c_{\mathcal{R}_{WC}^{MP}}(\omega)$	0	1	1	0	2	2	1	1
$\kappa^Z_{\mathcal{R}_{WC}^{MT}}(\omega)$	1	2	2	0	2	2	0	0
$\kappa^c_{\mathcal{R}_{WC}^{MT}}(\omega)$	1	2	2	0	2	2	0	0

Table 9 Inferences drawn from the formal systems for the knowledge bases of the Suppression Task

System	Realization							
	Plain		Connecting premise		Necessary condition		Weak completion	
	MP	MT	MP	MT	MP	MT	MP	MT
Reiter default logic	$^{n}_{s}$	s	s	s	s	s	s	s
System P	$^{n}_{s}$	$^{n}_{s}$	s	s	s	s	s	s
System Z	$^{n}_{s}$	$^{n}_{s}$	s	s	s	s	s	s
c-Representations	$^{n}_{s}$	$^{n}_{s}$	s	s	s	s	s	s

Here, $^{n}_{s}$ indicates that the system draws the conclusion according to the formal rule (modus ponens or modus tollens) and does not show the suppression effect and s indicates that this inference cannot be drawn, that is, the suppression effect occurs. The definitions for the different knowledge bases can be found in Tables 5 and 6

5 Predictions of the Formal Systems When Weakening Strict Knowledge into Defeasible Knowledge

In the previous section, a potential defeater for inference using modus ponens or modus tollens in the form of the additional statement "if the library is open, she will study late in the library", that is, an additional prerequisite for "her being in the library", has been added explicitly to the conditional knowledge base. Nonetheless, even if such a defeater is not added explicitly, the potential existence of exceptions for the rules in the knowledge base might weaken the MP or MT inference. In this section we examine the systematical weakening of knowledge in the knowledge base. We start with purely strict knowledge of a rule and a fact and examine, whether, and if, how the inferences change either from a strict fact to a defeasible fact, or from a strict rule to a defeasible rule, or both. Table 10 (leftmost column) lists these variations for the MP and MT cases, with a listing of the possible worlds for these knowledge bases (next four columns). We examine these variations of knowledge for the systems presented in Sect. 3 and classical propositional logic. The inferences drawn by the systems under the given variation to be compared to the inferences human reasoners draw given the same variations, the following section reports the experiments to determine these inferences.

Class 0: Propositional Logic
Propositional logic cannot express defeasible knowledge, which was the reason for looking into approaches of nonmonotonic reasoning, in the first place. For $\mathcal{R}^{MP}_{CC}$ and $\mathcal{R}^{MT}_{CC}$, propositional logic will license for the classical inferences l for modus ponens and $\overline{e}$ for modus tollens, since these are the classical inferences. For all other cases, this approach is not applicable.

Class 1: System P
As defined in Sect. 3.1, we can infer a formula ψ from a formula ϕ from knowledge base $\mathcal{R}$ iff the union of $\mathcal{R}$ with the conditional $(\overline{\psi}|\phi)$ is inconsistent. For $\mathcal{R}^{MP}_{CC}$ and $\mathcal{R}^{MT}_{CC}$ there is only one world possible, which falsifies $(\overline{l}|\top)$ (for the MP case) or $(e|\top)$ (for the MT case). Therefore we conclude that in this case the classical

Table 10 From left to right: knowledge bases varied between purely strict and purely conditional knowledge, possible worlds for these knowledge bases, the variable of interest for either the MP or MT case, the plausibility of the respective literals in the ranking approaches and the result of the inference in the ranking approaches

Knowledge base	Possible worlds el	$e\bar{l}$	$\bar{e}l$	$\bar{e}\bar{l}$	Variable of interest	Plausibility of Literals		Result
$\mathcal{R}_{CC}^{MP} = (\{e \Rightarrow l, e\}, \varnothing)$	✓				L	$\kappa(l) = 0$	$\kappa(\bar{l}) = \infty$	l
$\mathcal{R}_{CW}^{MP} = (\{e\}, \{(l\|e)\})$	✓	✓			L	$\kappa(l) = 0$	$\kappa(\bar{l}) = 1$	l
$\mathcal{R}_{FW}^{MP} = (\{e \Rightarrow l\}, \{(e\|\top)\})$	✓		✓	✓	L	$\kappa(l) = 0$	$\kappa(\bar{l}) = 1$	l
$\mathcal{R}_{CFW}^{MP} = (\varnothing, \{(l\|e), (e\|\top)\})$	✓	✓	✓	✓	L	$\kappa(l) = 0$	$\kappa(\bar{l}) = 1$	l
$\mathcal{R}_{CC}^{MT} = (\{e \Rightarrow l, \bar{l}\}, \varnothing)$				✓	E	$\kappa(e) = \infty$	$\kappa(\bar{e}) = 0$	$\bar{e}$
$\mathcal{R}_{CW}^{MT} = (\{\bar{l}\}, \{(l\|e)\})$		✓		✓	E			
$\mathcal{R}_{FW}^{MT} = (\{e \Rightarrow l\}, \{(\bar{l}\|\top)\})$	✓		✓	✓	E	$\kappa(e) = 1$	$\kappa(\bar{e}) = 0$	$\bar{e}$
$\mathcal{R}_{CFW}^{MT} = (\varnothing, \{(l\|e), (\bar{l}\|\top))$	✓	✓	✓	✓	E	$\kappa(e) = 1$	$\kappa(\bar{e}) = 0$	$\bar{e}$

Worlds that are possible in the respective realizations are denoted with ✓; worlds shaded gray are worlds that contradict the designated MP/MT inference. Since $\mathcal{R}_{CW}^{MT}$is inconsistent, there is neither a c-representation nor a System Z OCF, hence neither an inference possible

inference is drawn. In the case $\mathcal{R}_{CW}^{MT}$, the only worlds possible falsify the weakened conditional, therefore the knowledge base is inconsistent. This means that every extension of the knowledge base is inconsistent, also. Therefore we can infer $\bar{e}$, as expected for modus tollens, but also e, which is inconsistent to the prior inference. For all other cases, System P licenses for the designated conclusions, that is, l for modus ponens and $\bar{e}$ for modus tollens.

Class 2: Logic Programming

We implemented the different weakenings of the conditionals in logic programming. In the following we test the predictions of the weak completion semantics for the MP and MT cases. Within this approach a natural language conditional "if e then l" is best represented by a logical formula of the form $l \leftarrow e \wedge \neg ab$. How can we apply this idea to the following 4 cases $\mathcal{R}_{CC}^{MP}, \mathcal{R}_{CW}^{MP}, \mathcal{R}_{CW}^{MP}$, and $\mathcal{R}_{CW}^{MP}$? Everything depends on the representation of the conditionals. A conditional "if e then normally l" can be represented by the same form like before $l \leftarrow e \wedge \neg ab_1$. For normally e, we have $e \leftarrow \neg ab_2$. For none of these cases something abnormal is known, by the closed world assumption, we can assume that $ab_1, ab_2 \leftarrow \bot$. Based on these deliberations, we construct the programs given in Table 11.

So for the given programs, the weak completion semantics always licenses for the modus ponens resp. modus tollens inference. However, the programs are smaller for $\mathcal{R}_{CC}^{MP}, \mathcal{R}_{CW}^{MP}$ than $\mathcal{R}_{FW}^{MP}$ and $\mathcal{R}_{CFW}^{MP}$ and less operations are necessary during the weak completion procedure. This allows for a prediction of which inference is, from a pure operational perspective, drawn with less effort.

$$\mathcal{R}_{CC}^{MP} = \mathcal{R}_{CW}^{MP} > \mathcal{R}_{FW}^{MP} = \mathcal{R}_{CFW}^{MP}$$

For all of the knowledge bases we have only one model. For all MP cases, this model licenses for the MP inference, that is, "she will study late in the library", and "she does not have an essay to write" in all MT cases.

Table 11 Implementation of the different weakening possibilities in logic programming

Program	Least-model	Inference
$\mathcal{P}^{MP}_{CC} = \{l \leftarrow e, e \leftarrow \top\}$	$(\{e, l\}, \emptyset)$	l
$\mathcal{P}^{MP}_{CW} = \{l \leftarrow e \wedge \overline{ab_1}, e \leftarrow \top, ab_1 \leftarrow \bot\}$	$(\{e, l\}, \{ab_1\})$	l
$\mathcal{P}^{MP}_{FW} = \{l \leftarrow e, e \leftarrow \overline{ab_2}, ab_2 \leftarrow \bot\}$	$(\{e, l\}, \{ab_2\})$	l
$\mathcal{P}^{MP}_{CFW} = \{l \leftarrow e \wedge \overline{ab_1}, e \leftarrow \overline{ab_2}, ab_1 \leftarrow \bot, ab_2 \leftarrow \bot\}$	$(\{e, l\}, \{ab_1, ab_2\})$	l
$\mathcal{P}^{MT}_{CC} = \{l \leftarrow e, l \leftarrow \bot\}$	$(\emptyset, \{e, l\})$	$\overline{e}$
$\mathcal{P}^{MT}_{CW} = \{l \leftarrow e \wedge \overline{ab_1}, l \leftarrow \bot, ab_1 \leftarrow \bot\}$	$(\emptyset, \{e, l, ab_1\})$	$\overline{e}$
$\mathcal{P}^{MT}_{FW} = \{l \leftarrow e, l \leftarrow ab_3, ab_3 \leftarrow \bot\}$	$(\emptyset, \{e, l, ab_3\})$	$\overline{e}$
$\mathcal{P}^{MT}_{CFW} = \{l \leftarrow e \wedge \overline{ab_1}, l \leftarrow ab_3, ab_1 \leftarrow \bot, ab_3 \leftarrow \bot\}$	$(\emptyset, \{e, l, ab_1, ab_3\})$	$\overline{e}$

Class 3: Reiter Default Logic

Applying Reiter Default Logic to the respective realizations of the weakening cases given in Table 12 we obtain that for each knowledge base we have exactly one extension. In each of the MP cases, this extensions includes l, so Reiter default logic licenses for the MP inference for each of the different weakened cases. In the *CC* case there is no default in $\mathcal{D}^{MP}_{CC}$ and therefore we have $Cn(\mathcal{W}^{MP}_{CC}) = Cn(\{e \Rightarrow l, e\}) = Cn(\{e, l\})$ as the only extension. For all other cases the defaults are applicable and the extension is $Cn(\{e, l\})$. See Fig. 4 for an example of a process tree of these different cases. In the MT cases, there is no default in $\mathcal{D}^{MT}_{CC}$ and therefore we have the extension $Cn(\mathcal{W}^{MT}_{CC}) = Cn(\{e \Rightarrow l, \overline{l}\}) = Cn(\{\overline{e}, \overline{l}\})$. In the *FW* case the default is applicable, so that $\overline{l}$ is put into the set *In* and we have the same extension as before. So we can infer e in the *CC* and *FW* cases. In the other two cases the default is not applicable and we get the extension $Cn(\overline{l})$. So we cannot infer $\overline{e}$ in the *CW* and *CFW* cases, that is, whenever we weaken the conditional "if she has an essay to write, she will study late in the library" we lose the possibility of inferring using modus tollens, whereas in cases where none or only the fact is weakened, the MT inference is still valid in the example presented.

Table 12 Reiter default logic realization of the classical (CC) case as well as the weakened cases where the conditional (CW) the fact (FW) or both (CFW) premises are weakened

	Modus ponens	Modus tollens
CC	$\left(\mathcal{W}^{MP}_{CC} = \{e \Rightarrow l, e\}, \mathcal{D}^{MP}_{CC} = \emptyset\right)$	$\left(\mathcal{W}^{MT}_{CC} = \{e \Rightarrow l, \overline{l}\}, \mathcal{D}^{MT}_{CC} = \emptyset\right)$
CW	$\left(\mathcal{W}^{MP}_{CW} = \{e\}, \mathcal{D}^{MP}_{CW} = \left\{\frac{e : \overline{ab_1}}{l}\right\}\right)$	$\left(\mathcal{W}^{MT}_{CW} = \{\overline{l}\}, \mathcal{D}^{MT}_{CW} = \left\{\frac{e : \overline{ab_1}}{l}\right\}\right)$
FW	$\left(\mathcal{W}^{MP}_{FW} = \{e \Rightarrow l\}, \mathcal{D}^{MP}_{FW} = \left\{\frac{\top : e}{e}\right\}\right)$	$\left(\mathcal{W}^{MT}_{FW} = \{e \Rightarrow l\}, \mathcal{D}^{MT}_{FW} = \left\{\frac{\top : \overline{l}}{\overline{l}}\right\}\right)$
CFW	$\left(\mathcal{W}^{MP}_{CFW} = \emptyset, \mathcal{D}^{MP}_{CFW} = \left\{\frac{e : \overline{ab_1}}{l}, \frac{\top : e}{e}\right\}\right)$	$\left(\mathcal{W}^{MT}_{CFW} = \emptyset, \mathcal{D}^{MT}_{CFW} = \left\{\frac{e : \overline{ab_1}}{l}, \frac{\top : \overline{l}}{\overline{l}}\right\}\right)$

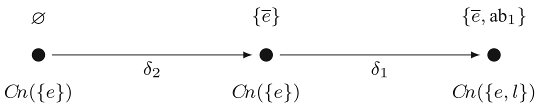

Fig. 4 The process tree visualizing the Reiter default processes for $(\mathcal{W}^{MP}_{CFW}, \mathcal{D}^{MP}_{CFW})$ where each vertex is annotated with the respective sets *In* (*below*) and *Out* (*above*)

Class 4: Ranking Models

For ranking models we invoke the inductive approaches System Z and c-representations on the conditional knowledge bases in Table 3 with the worlds possible given the respective strict formulas. For each of these knowledge bases, both approaches yield identical ranking functions. As already stated in the section regarding System P, $\mathcal{R}^{MT}_{CW}$ is inconsistent, therefore, Algorithm 1 cannot construct partitions, and the system of inequalities (6) is not solvable (cf. Pearl 1990; Kern-Isberner 2004). So we cannot infer anything from $\mathcal{R}^{MT}_{CW}$. Based on all other knowledge bases, the inference for l (in the MP case) and $\overline{e}$ (in the MT case) is possible: Here, even if there are worlds possible which are models of the inverse (i.e., $\overline{l}$ or e), each (most) plausible (that is, preferred) world is a model of the designated inference, as given in Table 10. In the MP case for the *CC* approach there is only one possible world *el*. This world is most plausible and the κ-value of l is 0. There is no world that satisfy $\overline{l}$ so we have $\kappa(\overline{l}) = \infty$. For all other approaches we have $\kappa(l) = 0 < 1 = \kappa(\overline{l})$ and therefore we can infer that "she will study late in the library". For the MT cases we have similar results. In the *CC* approach there is only one possible world $\overline{e}\overline{l}$ that does not satisfy e, therefore we have $\kappa(e) = \infty$. $\mathcal{R}^{MT}_{CW}$is inconsistent and there is no OCF possible. In the other two approaches we have $\kappa(\overline{e}) = 0 < 1 = \kappa(e)$ and hence "she has no essay to write".

Overview of Each System's Prediction

Table 13 summarizes the results of the capabilities of the systems and inferences given the weakened conditionals.

Table 13 Each system's prediction for the four cases CC, CW, FW, and CFW

Inference	CC		FW		CW		CFW	
System	MP	MT	MP	MT	MP	MT	MP	MT
Propositional logic	●	●						
Reiter default logic	●	●	●	●	●	–	●	–
System P	●	●	◐	◐	○	↯	○	○
c-Representations	●	●	◐	◐	○		○	○
System Z	●	●	◐	◐	○		○	○
Weak completion semantics	●	●	●	●	●	●	●	●

Please note that there are differences to the general case (e.g., Reiter does not infer MT in general cf. Table 9). Notation: ⟨empty cell⟩ not applicable; –: cannot be inferred; ○: can be inferred (i.e., holds in all preferred worlds/all extensions); ◐: can be inferred and holds in most possible worlds/extensions; ●: can be inferred and holds in all worlds possible/all extensions; ↯ inconsistent

6 Human Inferences in the Light of Defeasibility: Experiments

Human reasoning about uncertainty is connected with knowledge about the existence of exceptions. In contrast to the interpretation of conditionals by material implication, few conditional statements in everyday life draw such a strong connection. As mentioned in the introduction, most previous psychological experiments have tested problems with possible defeaters presented explicitly. In this section, we tested how reasoners deal with the implicit activation of the knowledge about exceptions through the insertion of keywords. An example in conditional reasoning is a structure like "if ...then normally ...". This does represent the idea of having a "license for inference" (Stenning and Lambalgen 2008), namely that a conditional such as "if e then l" is best represented by $l \leftarrow e \wedge \overline{ab_1}$, i.e., "if e and nothing *abnormal*$_1$ is known then l", or equivalently "if e then normally l". In this section, we investigate how human reasoners reason with the insertion of the nonmonotonic keyword "normally".

Our hypotheses of the nonmonotonic reasoning with an implicit trigger keyword are:

H1: The MP and MT inferences are drawn more cautiously.
H2: The MP and MT inferences are drawn slower.
H3: The "may or may not" conclusion is drawn more often.
H4: More uncertainty is triggered by the keyword "normally" when it is inserted in the premise (Table 1), which is a specific difference predicted by the systems (see Sect. 5 and Table 10).
H5: Defeasibility in conditionals leads to more cautious inferences than uncertainty in facts.

We briefly present the definitions for each of the four cases here:

	if ϕ then ψ	if ϕ then normally ψ
ϕ	*CC*	*CW*
Normally, ϕ	*FW*	*CFW*

From these hypotheses, we derived the following order of the keyword cases with respect to the cautiousness inferences are drawn:

$$CC \geq FW \geq CW \geq CFW$$

That is, we assume participants to endorse the modus ponens and modus tollens inferences more often and faster for the *CC* case and decrease to the most cautious inferences for *CFW*.

6.1 Experiment I: Interpretation of Nonmonotonic Keywords

Uncertainty can be triggered by different keywords such as "most", "normally", "often", etc. But how do participants interpret such quantifiers numerically? We were interested in the amount of possible exceptions they may implicitly with the presence of such keyword.

6.1.1 Participants

Eighty-one participants ($M = 33.6$ years, 24f) participated in an internet study on Amazon Mechanical Turk. We took the general precautions about native speakers of English. They were paid for their participation.

6.1.2 Design and Procedure

All participants received, in a randomized order, 14 statements consisting of one of the nonmonotonic keywords: “all”, “usually”, “most”, “the majority of”, “few”, “seldom”, “normally”, “typically”, “often”, “some”, “the minority of”, “hardly”, “rarely”, and “no”. For example, for “typically”, the statement is like “Typically, objects of type A are also of type B”. Then, the question “How many percent of objects of type A are also of type B?” was presented. The task of the participants was to give the lower and upper bounds for their numerical interpretation of the respective keywords both on the scale from 0% to 100%, respectively. They could use as much time as they wanted for the valuations.

6.1.3 Results

A summary of the valuations can be found in Fig. 5. We calculated the median of the min- and max-values as the interval frontiers. It becomes obvious that different keywords can trigger different ranges for the amount of exceptions. The keyword “normally” allows for some exceptions, but not all (no participant chose 100 as the maximum). The range between the medians of the max- and min-values spans the interval from 90% to 60%.

6.2 Experiment II: Defeasible Conditional Reasoning

In this experiment, we systematically varied the strictness and defeasibility for the relations and tested the predictions in Table 10.

6.2.1 Participants

We tested 119 participants ($M = 38.6$ years, 51f) in an internet study on Amazon Mechanical Turk from a different sample than the two previous experiment. They were paid for their participation. 20 participants were excluded from data analyses due to missing responses.

6.2.2 Design and Procedure

Each participant has been randomly assigned to one of the four groups (cp. Table 1). Participants had to respond to 16 problems in two blocks: In the first block, the first group received the strict conditional problems (without a “normally”), the second group received a “normally” in the conditionals (i.e., $\mathcal{R}_{CW}^{MP}$ and $\mathcal{R}_{CW}^{MT}$), the third

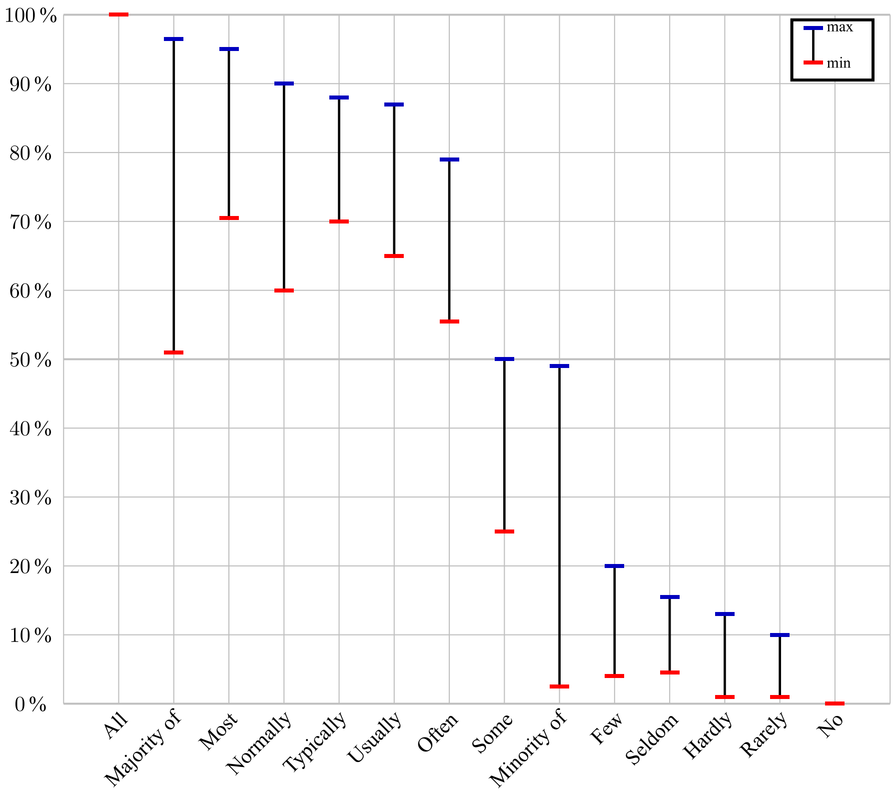

Fig. 5 The interpretation intervals of nonmonotonic keywords. The *red* and *blue caps* represent the respective medians of the min- and max-values participants gave in their responses

group received strict conditionals and a "normally" in the facts (i.e., $\mathcal{R}^{MP}_{FW}$ and $\mathcal{R}^{MT}_{FW}$), and the fourth group received a "normally" in the conditionals and in the facts (i.e., $\mathcal{R}^{MP}_{CFW}$ and $\mathcal{R}^{MT}_{CFW}$). Each of the four inference forms (MP, MT, DA, AC) was presented twice with different content, hence, in Block I, each participant had to solve eight problems. The AC and DA problems were used as fillers. After the first block, participants of the first and fourth group received 8 problems in the fourth (i.e., $\mathcal{R}^{MP}_{CFW}$ and $\mathcal{R}^{MT}_{CFW}$) and first case (i.e., the classical conditionals) respectively. Participants that received problems of the second (i.e., $\mathcal{R}^{MP}_{CW}$ and $\mathcal{R}^{MT}_{CW}$) and third case (i.e., $\mathcal{R}^{MP}_{FW}$ and $\mathcal{R}^{MT}_{FW}$) in the first block received in the second block then the third and second case respectively.The table below shows in detail how the blocks of different keyword cases were presented to the four groups of participants.

	Group I	Group II	Group III	Group IV
Block I	*CC*	*CW*	*FW*	*CFW*
Block II	*CFW*	*FW*	*CW*	*CC*

6.2.3 Results

The overall correctness for the classical MP (92%) and the MT (86%) were both high and comparable to previous findings. We compared the percentages of correct response and median response times according to the problem type and keyword cases (8 groups). No difference could be found for the percentage of correct responses and response times between Block I and Block II, except for the percentage of correct responses for the MT -type of case *CW* (Wilcoxon-Signed-Rank: $z = 2.236,\ p < .05$).

Therefore, we aggregated the results of the two blocks in the following analysis, except for the Jonckheere–Terpstra test. We compared all the four keyword cases wrt. MP and MT inferences drawn. The number of MP and MT inferences can show if our hypotheses H1 and H3 hold. The median response times are expected to show the pattern as suggested in our hypothesis H2.

The results support our hypothesis H1 that when defeasibility is introduced in a knowledge base, the reasoners retract from drawing MP and MT inference and our hypothesis H3 that participants produce more the cautious "may or may not" conclusion with defeasibility applied in a conditional reasoning problem. However, the results did not support our hypothesis H5, especially for the MP inferences. It seems that participants had similar performance no matter the defeasibility is in the fact or in the conditional (Table 14).

For the data analysis of response time, we have excluded response times of incorrect responses, i.e., only the response times for the specific inference drawn were included:

		CC (s)	*FW* (s)	*CW* (s)	*CFW* (s)
MP	Median	5.81	6.70	6.23	8.34
	Median absolute deviation	2.73	2.87	2.79	3.87
MT	Median	7.57	8.14	6.56	8.99
	Median absolute deviation	3.65	3.06	3.30	4.09

The Jonckheere–Terpstra test for Block I data showed that there was a reliable trend of lower percentage of correct responses with more defeasibility applied in the conditionals (from CC, CW, FW to CFW, $T_{JT} =$ 26,626, $z = 2.636, p < .01$) and a

Table 14 The results of Experiment II: The rounded percentage each group of participants applied the inference rules MP and MT

	Response	*CC*	*FW*	*CW*	*CFW*
MP	Correct response	92	82	67	69
	"May or may not"	8	18	31	31
MT	Correct response	86	63	71	62
	"May or may not"	10	31	29	37

For the MP inferences, the differences between *CC* and *FW* to *CW* and *CFW* were significant ($p < .01$). For the MT inferences the differences between *CC* and the other 3 groups were significant ($p < .01$)

significant slower response time trend in the same order, $T_{JT} = 35{,}956$, $z = 4.958, p < .01$, when we aggregated the results of MP and MT inferences.

6.3 Experiment III: Stability of Nonmonotonic Inferences

In this experiment, we varied the strictness and defeasibility for the relations as in Experiment II, however, the problems were presented in a randomized order, instead of in a block design. The aim of this experiment was to test if there is a learning effect in the participants after solving 8 inference problems of different forms. The second question was whether solving a problem in a specific case can directly influence the inference of problems in another case.

6.3.1 Participants

We received data from 73 participants ($M = 35.8$ years, 58f) in an internet study on Amazon Mechanical Turk from a different sample than the two previous experiments. They were paid for their participation.

6.3.2 Design and Procedure

Each participant solved 16 conditionals with different contents, which were randomly assigned. Participants have been randomly assigned to one of the two groups: The first group received the four strict conditional reasoning problems (i.e., MP, MT, DA, and AC; without a “normally”) and the four modified problems with a ”normally” in the conditional and in the fact (i.e., $\mathcal{R}^{MP}_{CFW}$ and $\mathcal{R}^{MT}_{CFW}$), each in two different contents in two blocks in a randomized order. The second group received the four problem types with a ”normally” in the conditional, but without a “normally” in the fact (i.e., $\mathcal{R}^{MP}_{CW}$ and $\mathcal{R}^{MT}_{CW}$), and the four classical conditionals with a ”normally” in the fact (i.e., $\mathcal{R}^{MP}_{FW}$ and $\mathcal{R}^{MT}_{FW}$, cp. Table 1). The DA and AC problems were included as fillers. After the first block of the eight problems, participants were tested with the same problem set but in a different content and randomized order. We tested our hypotheses H1 to H3 with Group I participants and H5 with Group II participants. In addition, by comparing the results of Block I and Block II, we can test whether there was any learning effect for solving this kind of problems.

6.3.3 Results

The overall percentage of correct responses for the classical MP (99%) and MT (89%) were both high and comparable to Experiment II. We compared the percentage of correct responses for Block I and Block II data with Wilcoxon-Signed-Rank test regarding participant group, keyword case, and problem types, and found no reliable difference. Therefore, we found no evidence supporting any learning effect during the task. As in Experiment II, we aggregated the results of the two blocks in the following data analyses. Furthermore, we compared all the four

Table 15 The results of Experiment III: the table presents the rounded percentage each group of participants applied the respective inference rules MP or MT or chose the “may or may not” response

	Cases	*CC*	*FW*	*CW*	*CFW*
MP	Correct response	99	80	74	61
	“May or may not”	0	20	26	39
MT	Correct response	89	53	64	59
	“May or may not”	7	37	31	37

For MP, the differences between *CC* and the other 3 cases are significant (CW: $z = 4.243, FW$: $z = 3.742, CFW$: $z = 5.385, p < .01$), with the Wilcoxon-Signed-Rank test; and the difference between *FW* and *CFW* is significant ($z = 2.402, p < .05$) as well. For MT, the differences between Classical and the other keyword cases are significant (CW: $z = 3.053, FW$: $z = 4.226, CFW$: $z = 4.158$, respectively, $p < .01$)

cases with respect to the MP and MT inferences drawn. The results are all comparable to those of Experiment II (Table 15).

For the data analysis of response times, like Experiment II, we have included only the response times of the specific inferences drawn only. The response time of *CC* was reliably faster than those of *CW* and *CFW* (Wilcoxon-Signed-Rank: $z = 2.158, p < .05$ and $z = 3.349, p < .01$, respectively); and that of *CW* was significantly faster than that of *CFW* (Wilcoxon-Signed-Rank: $z = 2.267, p < .05$), for MP inferences. While for MT-type, the response time of *CW* was reliably faster than that of *FW* (Wilcoxon-Signed-Rank: $z = 1.984$, $p < .05$).

		CC (s)	*FW* (s)	*CW* (s)	*CFW* (s)
MP	Median	5.63	6.81	7.05	7.93
	Median absolute deviation	2.09	2.71	2.94	2.91
MT	Median	7.26	9.00	8.07	8.52
	Median absolute deviation	2.55	5.22	2.42	2.58

Group I participants drew the MP inferences more often in the *CC*-case than in the *CFW*-case (Wilcoxon-Signed-Rank, $z = 6.604, p < .01$). Group II participants made significantly more MP than MT inferences ($z = 3.108, p < .01$). For response times, there is a reliable difference between both the keyword case and problem type for both Group I and Group II participants: For Group I participants, *CC* problems were solved faster than *CFW* ($z = 3.190, p < .01$), and MP are faster than MT inferences (Wilcoxon-Signed-Rank: $z = 2.371, p < .05$). Group II participants, drew *FW* problems faster than CW (Wilcoxon-Signed-Rank: $z = 2.010, p < .05$), and MP faster than MT inferences (Wilcoxon-Signed-Rank: $z = 2.915, p < .01$).

6.4 Theory-Evaluation

In this section, we briefly compare the results from the behavioral experiments with the predictions of the systems. The findings from the previous experiments indicate a similar trend between some systems (c-representations, System P, and System Z) and the behavioral results (see Page’s trend test above).

6.4.1 Do Systems Draw the Same Inferences as the Majority of Human Reasoners?

If we compare the predictions of the systems (Table 13) with the behavioral results, it is obvious that propositional logic is not suited to explain reasoning with uncertainty (as it provides predictions only for the case of classical conditional inferences, i.e., only in 2/8 cases tested here). Another interesting observation is that none of the systems, except the weak completion semantics (WCS), derive a MT inference for *CW* which has been drawn by about 71% (in Experiment II) and 64% (in Experiment III) of the participants.

6.4.2 Secondly, Can the Trend of the Data be Predicted by the Systems?

Three of the systems make ordered predictions based on the problems. The Jonckhere-Trend test showed that the predictions of the three systems are reliable (in the order: $CC \geq FW >$ CW $\geq CFW$) for MP inferences, $\mathrm{T_{JT}} = 6286, z = 3.038, p < .01$ and $\mathrm{T_{JT}} = 9404, z = 4.622, p < .01$ for percentage of correct responses and response time, respectively. We have excluded *CW* in the analyses for MT inferences (label as MT') as this keyword case is either inconsistent or not applicable for most, except one, systems: $\mathrm{T_{JT}} = 3374, z = 1.894, p = .058$ and $\mathrm{T_{JT}} = 4468, z = 2.177, p < .05$ for correctness of responses and response time, respectively. The Jonckhere-Trend test shows that the predictions of the three system are marginally reliable for percentage of correct responses and reliable for response times in general.

Additionally, we generated multinomial process tree-like structures (Batchelder and Riefer 1999) based on their predictions for different reasoning systems (cp. Table 16). The results show that Sysc* (aggregating c-representations, System P, and System Z; as they generate the same predictions for MP and MT') makes the

Table 16 Results of the multinomial process tree (MPT) analysis for trees implementing the different systems predictions

	G^2	FIA	BIC	AIC
MP				
Reiter default logic	29.978	27.319	59.886	39.978
Sysc*	**3.354**	17.569	**45.224**	**17.354**
WCS	29.978	27.319	59.886	39.978
WCS$^+$	**3.354**	**16.873**	51.205	19.354
MT				
Reiter default logic	**2.024**	**8.234**	**17.889**	**8.024**
Sysc*	2.567	133.687	36.908	14.56
WCS	12.406	18.529	42.313	22.406
WCS$^+$	4.879	17.648	52.731	20.879

Sysc* represents c-representations and the Systems P and Z as they generate the same predictions. We have excluded *CW* for MT inferences as only one system generates a prediction. WCS$^+$ is the weak completion semantics (WCS) extended with the number of operations to derive the inference. The best results for each measure are given in bold script best predictions for MP based on the AIC and BIC values. However, WCS^+ is the best based on the FIA value (Singmann and Kellen 2013). For the reduced MT' cases (without the *CW*-case), Reiter Default Logic makes the best predictions according to the AIC, BIC, and FIA values. This finding is remarkable as Reiter Default Logic in general does not allow to draw a MT inference.

If we consider additional effects, e.g., the number of operations in WCS, then the performances of the system prediction improve as indicated in WCS^+ and WCS^+ is comparable to Sysc*. However, no one has proposed such a complexity measure so far.

7 Summary and General Discussion

Classical propositional logic and especially material implication are not suitable for explaining the peculiarities of human conditional reasoning (Wason 1968; Klauer et al. 2007). In addition, it is questionable if human reasoning is *monotonic*, i.e., the trueness of previously drawn inferences will not be affected by the acquisition of new knowledge. New information can lead to cautious inferences and human reasoners do not demonstrate to have a fix form of rules they have just applied, as shown by the context dependency of formally equivalent problems (Klauer et al. 2007). Hence, it is possible that human reasoners apply modus ponens and modus tollens or not depends on the scenario. This is a feature that humans share with many formal nonmonotonic inference relations (cp. Sect. 3).

Formal systems generate clear predictions which are implemented and thus ready-to-use. They are characterized with respect to their formal and computational properties. In this study, we have done the analyses based on de-facto standards, such as Reiter Default Logic, System P, System Z, representations, and Logical Programming with Weak Completion Semantics. Different nonmonotonic classes are covered in these systems. Formal nonmonotonic systems do not license for MT inferences as a general rule. Even more, if modus tollens would hold generally, this would render the system to be a system of monotonic inference Kraus et al. (1990).

Our findings in the experiments support inference by preferential models, i.e., the inference by System P and the ranking approaches, to be able to capture the semantic difference in the Suppression Task which humans demonstrate but only if explicit and implicit knowledge are modeled carefully and thoroughly. In these models, ϕ preferentially entails ψ if and only if ψ holds in the most plausible models of ϕ. Figure 6 illustrates how the models of the two formulas ϕ and ψ can be related to each other if $\phi \mid\!\sim \psi$ is assumed. It can be seen that even if ψ holds in the most plausible models of ϕ, there are most plausible models of $\overline{\psi}$ which are models of ϕ and others which are models of $\overline{\phi}$, and hence no inference with respect to the outcome of ϕ (and thus the application of modus tollens) is possible. However, it is reasonable to draw the MP conclusion. That means, if the agent is in a situation where ϕ holds, it is most plausible that ψ is also true.

As a result of the investigation of the formal systems, we have shown that the suppression effect can be modeled in all the approaches of nonmonotonic reasoning

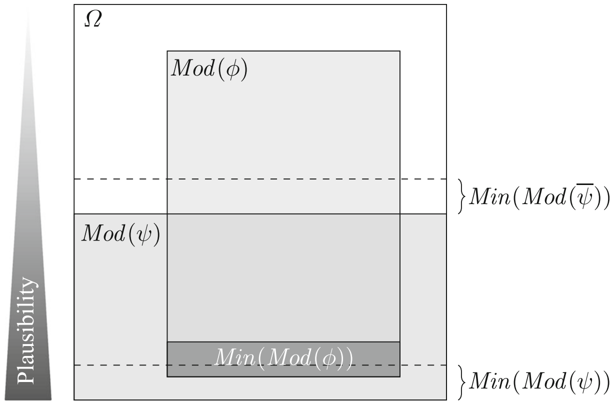

Fig. 6 Illustration of possible plausibility arrangement of models of formulas ϕ and ψ if $\phi \mid\!\sim \psi$

presented. It requires, however, different additional efforts to do so. Some approaches require to take hidden assumptions into account, others to represent the background knowledge explicitly and include it in the respective knowledge base. This finding should not be interpreted in a way such that the goodness-to-fit of the approaches depend on a skillful implementation alone. The manipulation is not for achieving our expected results, but to sharpen our senses in the search for an explanation of the peculiarities of the human reasoning system.

The presented systems are formal ones and deal with, and only with, the given information in a formal way. Instances of these variables are in principle unconnected and do not contain meaning of any type, other than the one formally specified[2]. For instance, in the running example, a human reasoner might—or most probably will—have a connection between, “being in a library” and “a library being open” in mind, and this connection might be a more general one that, for instance, states the general situation “one cannot be in a public institution if the building is closed”[3] This connection is not made in most of the plain approaches presented in Sect. 3, here, only Logical Programming connects the variables e and o in a rules body via the equivalence induced by the Weak Completion Semantics and the abnormality predicates.

To evaluate the predictive power of the formal nonmonotonic systems, we have tested a modeling of the MP inference in the Suppression Task (Byrne 1989) in a previous article Ragni et al. (2016). The work is extended in this study. Firstly, we included predictions for the MT case as well, apart from only the MP case in the first

[2] For instance, it is without problem to introduce variables F and M to a logical system and, without stating any conditions that these may be mutually exclusive, have states of worlds or individuals for which, one, both, or neither of these variables are true, despite the fact that the user of the formal system intended to encode individuals to *female* or *male* with these variables.

[3] That is, “*in general* one cannot be...”. One might have a key, have been (by accident or deliberately) locked in, have broken in, etc. But in the plausible situations, this rule and thus this connection between the variables holds.

modeling. In the second step, we investigated what additional (or background) information a knowledge base needs to contain, so that the systems which deviate from human defeasible inferences in the Suppression Task, originally, show the suppression effect. These modeling results demonstrate that human reasoners might assume additional background knowledge implicitly or diverse understanding of the premise information when reasoning from these premises. This needs to be investigated in future studies. In the third step, we generated fine-grained predictions of the different nonmonotonic systems for the nonmonotonic keyword "normally" (cp. Table 13). These predictions allow us to test the cognitive adequacy of these models in the final step.

Previous empirical approaches have not analyzed the variations of these nonmonotonic keywords systematically. While (Politzer 2005) used the meta-statement that the following assertion is "not certain". We were interested in using positive keywords that may trigger a reasoner to consider exceptions of the present conditional implicitly, without the need to think about a specific number of exceptions as in some other studies. Consequently, we first analyzed the interpretation of different keywords. Among all the possible keywords we tested, "normally" showed a distribution with at least over 60% of the cases being evidence for the conditional statement, but it allows for some exceptions in the representation as well. A byproduct is that it coincides with the formulation of a conditional with an abnormality predicate $l \leftarrow e \wedge \neg ab$, or in other words with the notion of "if e then normally l". Hence, insertion of the keyword "normally" can admit some abnormalities.

In the second experiment, we investigated the effect of the nonmonotonic keyword "normally" on drawing the modus ponens or modus tollens and if there is any differences between its insertion in the conditional or in the fact, or in both. The results supports several of our hypotheses. First of all, it triggered more cautious inferences, especially when the keyword is present in the conditional. This has been predicted by several nonmonotonic systems such as c-representations, System Z and System P. In this sense, the rule dominates a fact in the modus ponens.

Also it strongly hints at a plausibility ordering of the possible worlds as predicted by Spohn's ranking functions (Spohn 2012). Such an ordering might be connected to a preferred mental model theory (Ragni and Knauff 2013). Such ranks (and the inductive methods to acquire them, like c-representations,) require additional empirical investigations. Secondly, the number of "may or may not" responses increased accordingly, as well as the response times. Hence, the uncertainty manipulation showed a clear trend effect. Additionally, it seems possible to trigger different "degrees" of nonmonotonicity, i.e., the more uncertainty is introduced, the more often the reasoner retract from drawing the MP inference. Participants received conditional reasoning problems of all the possible inferences types, i.e., MP, MT, DA, and AC, with the same keyword case within a block. That means, the insertion of "normally" or not in the conditional or fact is identical for all the four problems. This block design was chosen to reduce the attention the keyword might draw. Participant may compare the problems with problem with other keyword cases when they are presented together. Furthermore, the high number of modus ponens (and modus tollens) inferences drawn despite the keyword "normally" draws a connection to the interpretation of Experiment I—that although there could be "abnormalities" these set are regularly false. Future work is required to investigate if any additional implicit assumptions exist in the completion process (instead of a weak completion) and the reasoning steps humans may draw.

The more the uncertainty is inserted, the less this abnormality is set to be false by individual reasoners. This prediction on an individual level is a strength of the logical systems in contrast to some probabilistic approaches that model the group decisions as a whole. Nonetheless, this aspect requires future research.

The third experiment investigated a group that received *CC* and *CFW* and a second group that received problems of keyword cases *FW* and *CW* randomly which were not presented in two different blocks. Such an "exposition" to conditionals or facts in strict and defeasible form should trigger a different behavior if participants would make the aforementioned comparison. Our analysis, however, showed that there were no differences from the results of Experiment II, nor between the first or the second half of the experiment according to different keyword cases. This indicates a form of stability in the inference process. And this stability, namely that the same inferences are drawn across time, is a common characteristic of formal systems. These findings show that human reasoners have not been influenced by this repeated presentation of similar problems.

The lower frequency of drawing a MT inference than a MP inference is consistent with other findings in the literature (Klauer et al. 2007). In both cases, the caution to draw the respective inference increased with the introduction of a nonmonotonic keyword like "normally". A difference between them is that the order in modus ponens of *CW* and *FW* is reversed. This might be related to the reasoning direction, being forward or backward, in a conditional inference (e.g., Klauer et al. 2007). In this article we focused on introduction of new information that supports or defeats knowledge, but there can be stronger candidates, namely *disablers* and *enablers* for conditionals. They can inhibit the application of a possible rule in the legal context (Gazzo Castañeda and Knauff 2016). Future work for the difference between such disablers and defeaters is required.

We put additional effort onto the question about how qualitative system predictions can be best evaluated. Multinomial process trees (MPTs) have been proven to be a successful method to model system predictions that incorporate some structural information (Ragni et al. 2014). Klauer et al. (2015) showed that ordered predictions for a task can be implemented in a MPT modeling. For our modeling approach, we had to represent predictions between different problem classes by MPTs. Future research is required for the potentials and limitations of MPT modeling. For our results, however, the evaluation of the theories show a rare congruence between all the relevant measures such as AIC, BIC, and FIA. If a complexity measure is introduced to the Weak Completion Semantics by counting the number of operations, then the Weak Completion Semantics reaches a similar performance like the three systems (c-representations, System P and System Z). While the Weak Completion Semantics focuses only on the minimal model, it remains an open question if ranks can be introduced as well within this system.

The knowledge bases presented in Sects. 4 and 5 are canonical models with respect to the implementation in the existing literature concerning psychological studies and their formal interpretation or normative predictions. Nonetheless, this may encode the knowledge in a different way from how it was intended:

Representing $(l|e)$ and $(e|\top)$ on the same level is (with exceptions) generally valid, as elements of the (background) knowledge of the agent yield not only the rule "if she has an essay to write, she will normally study in the library" but also the rule that "(usually/most of the time/normally) she has an essay to write".

This may not meet the intuition. Therefore, we proposed to introduce an element of dynamics into the discussion and argued, in the sense of Ramsey (1929), that an inference should indeed be a (hypothetical) revision of the epistemic state of an agent. By modeling reasoning in this way, we split up the epistemic state, which we set based on the background knowledge, and the inference by (virtually) revising this epistemic state with the premise of the conditional statement to be tested. And with this approach, we distinguish between knowledge of the agent and queries that are only relevant for the inference but have no impact on the knowledge of the agent. Thus, the model is able to reason not only about true events but also hypothetically or counterfactually. Usually, we do not expect her to have an essay to write, but if she had an essay to write, then the agent could infer whether she would be in the library or not. For this paper, we decided to keep the formal modeling as simple and uniform as possible and stick to the static modeling because revision of the logic programs under the Weak Completion Semantics and Reiter default logic has hardly been ever considered. Moreover, for the ranking models, the same beliefs would obtain in this simple example for both the static and dynamic modeling.

These findings can help as well to build better formal commonsense reasoning systems that can take typical inferences of humans into account. And thus, such systems will be more successful in interacting with humans. Finally, the results showed that to distinguish between a "correct" and an "incorrect" inference in many reasoning problems requires at least two additional information: The first being the type of inference systems used and the second being the kind of the elements of the knowledge base. And one extra information, probably being the most important one, is: Do you know any exceptions to the rule? These three points have indicate why classical logic is not sufficient to model human reasoning and why nonmonotonic systems that take the "right" knowledge base and knowledge about possible exceptions into account with an presupposed ordering on models are superior to other systems.

Acknowledgements This work is supported by DFG-Grants KI1413/5-1 to G. Kern-Isberner and by RA1934 2/1, RA 1934 4/1 and Heisenberg DFG fellowship RA1934 3/1 to M. Ragni. T. Bock and C. Eichhorn are supported by Grant KI1413/5-1 and A. P. P. Tse by RA1934 2/1. The authors would like to thank Richard Niland and Daniel Lux for discussions.

References

Adams, E. (1965). The logic of conditionals. *Inquiry*, *8*(1–4), 166–197.

Antoniou, G. (1997). *Nonmonotonic reasoning*. Cambridge, MA: MIT Press.

Baratgin, J., Over, D., & Politzer, G. (2014). New psychological paradigm for conditionals and general de finetti tables. *Mind & Language*, *29*(1), 73–84.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86.
Beierle, C., Eichhorn, C., & Kern-Isberner, G. (2016). Sceptical Inference based on C-representation and its characterization as a constraint satisfaction problem. In *Proceedings of the 9th International Symposium on Foundations of Information and Knowledge Systems (FoIKS 2016), Lecture Notes of Computer Science* (Vol. 9616, pp. 65–82). Berlin, DE: Springer.
Beierle, C., & Kern-Isberner, G. (2014). *Methoden wissensbasierter Systeme, 5. überarbeitete und erweiterte Auflage*. Wiesbaden, DE: Springer Vieweg **(in German)**.
Bochman, A. (2001). *A logical theory of nonmonotonic inference and belief change*. Berlin, DE: Springer.
Bonnefon, J. F., Da Silva Neves, R., Dubois, D., & Prade, H. (2006). Background default knowledge and causality ascriptions. In: G. Brewka, S. Coradeschi, A. Perini, & P. Traverso (Eds.), *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 2006), Frontiers in Artificial Intelligence and Applications* (Vol. 141, pp. 11–15). Amsterdam, NL: IOS Press.
Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 61–83.
Byrne, R. M. (1991). Can valid inferences be suppressed? *Cognition*, *39*(1), 71–78.
Byrne, R. M., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*(3), 347–373.
Da Silva Neves, R., Bonnefon, J. F., & Raufaste, E. (2002). An empirical test of patterns for nonmonotonic inference. *Annals of Mathematics and Artificial Intelligence*, *34*(1–3), 107–130.
Dietz, E., Hölldobler, S., & Ragni, M. (2012). A computational approach to the suppression task. In N. Miyake, D. Peebles, & R. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1500–1505). Austin, TX: Cognitive Science Society.
Dietz, E. A., & Hölldobler, S. (2015). A new computational logic approach to reason with conditionals. In: *International conference on logic programming and nonmonotonic reasoning* (pp. 265–278). New York: Springer.
Dubois, D., & Prade, H. (1996). Conditional objects as nonmonotonicConsequence relations. In *Principles of knowledge representationand reasoning: Proceedings of the 4th international conference (KR'94)* (pp. 170–177). San Francisco, CA: Morgan Kaufmann Publishers.
Dubois, D., & Prade, H. (2015). *Possibility theory and its applications: Where do we stand?*. Berlin: Springer.
Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, *77*(2), 321–357.
de Finetti, B. (1974). *Theory of probability* (Vol. 1,2). New York, NY: Wiley.
Fitting, M. (1985). A Kripke–Kleene semantics for logic programs. *Journal of Logic Programming*, *2*(4), 295–312.
Gabbay, D. (1985). Theoretical foundations for non-monotonic reasoning in expert systems. In K. R. Apt (Ed.), *Logics and models of concurrent systems* (pp. 439–457). New York, NY: Springer.
García, A. J., & Simari, G. R. (2004). Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, *4*, 95–138.
Gazzo Castañeda, E. L., & Knauff, M. (2016). Defeasible reasoning with legal conditionals. *Memory & Cognition*, *44*(3), 499–517. doi:10.3758/s13421-015-0574-7.
Goldszmidt, M., & Pearl, J. (1996). Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, *84*(1–2), 57–112.
Halpern, J. Y. (2005). *Reasoning about uncertainty*. Cambridge, MA: MIT Press.
Hölldobler, S., & Kencana Ramli, C. D. (2009a). Logic programs under three-valued Łukasiewicz semantics. In P. M. Hill, & D. S. Warren (Eds.), *Logic programming, 25th international conference, ICLP 2009, LNCS* (Vol. 5649, pp. 464–478). Heidelberg: Springer.
Hölldobler, S., & Kencana Ramli, C. D. (2009b). Logics and networks for human reasoning. In C. Alippi, M. M. Polycarpou, C. G.Panayiotou, & G. Ellinas (Eds.), *International conference on artificial neural networks, ICANN 2009, part II, LNCS* (Vol. 5769, pp. 85–94). Heidelberg: Springer.
Johnson-Laird, P., & Byrne, R. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646–677.
Kern-Isberner, G. (2001). *Conditionals in nonmonotonic reasoning and belief revision*. No. 2087 in LNCS. Berlin, DE: Springer.
Kern-Isberner, G. (2004). A thorough axiomatization of a principle of conditional preservation in belief revision. *Annals of Mathematics and Artificial Intelligence*, *40*, 127–164.
Kern-Isberner, G., & Eichhorn, C. (2012). A structural base for conditional reasoning. In: *Human reasoning and automated deduction—KI 2012 workshop proceedings* (pp. 25–32).

Klauer, K. C., Singmann, H., & Kellen, D. (2015). Parametric order constraints in multinomial processing tree models: An extension of Knapp and Batchelder (2004). *Journal of Mathematical Psychology*, *64*, 1–7.

Klauer, K. C., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 680–703.

Kleene, S. C. (1952). *Introduction to metamathematics*. Amsterdam: Bibl. Matematica.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*(1–2), 167–207.

Kripke, S. A. (1972). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Kuhnmünch, G., & Ragni, M. (2014). Can formal non-monotonic systems properly describe human reasoning? In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1222–1227). Austin, TX: Cognitive Science Society.

Lehmann, D. J., & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, *55*(1), 1–60.

Lewis, D. K. (1986). *On the plurality of worlds*. Hoboken, NJ: Blackwell Publishers.

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

Łukasiewicz, T. (2005). Weak nonmonotonic probabilistic logics. *Artificial Intelligence*, *168*(1–2), 119–161.

Makinson, D. (1994). General patterns in nonmonotonic reasoning. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming* (Vol. 3, pp. 35–110). New York, NY: Oxford University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., & Chater, N. (2016). Probabilities, causation, and logic programming in conditional reasoning: Reply to Stenning and van Lambalgen. *Thinking & Reasoning*, *22*(3), 336–354.

Pearl, J. (1990). System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh (Ed.), *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge (TARK1990)* (pp. 121–135). San Francisco, CA: Morgan Kaufmann Publishers Inc.

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

Politzer, G. (2005). Uncertainty and the suppression of inferences. *Thinking & Reasoning*, *11*(1), 5–33.

Ragni, M., Eichhorn, C., & Kern-Isberner, G. (2016). Simulating human inferences in the light of new information: A formal analysis. In S. Kambhampati (Ed.). *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)* (pp. 2604–2610).

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*(3), 561–588.

Ragni, M., Singmann, H., & Steinlein, E. M. (2014). Theory comparison for generalized quantifiers. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1330–1335). Austin, TX: Cognitive Science Society.

Ramsey, F. P. (1929). General propositions and causality. In *Philosophical papers* (pp. 145–163). Cambridge, UK: Cambridge University Press.

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

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560–575.

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26–36.

Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in decision, belief change and statistics: Proceedings of the Irvine Conference on Probability and Causation, the Western Ontario Series in Philosophy of Science* (Vol. 42, pp. 105–134). Dordrecht, NL: Springer.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford, UK: Oxford University Press.

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

Thorn, P. D., Eichhorn, C., Kern-Isberner, G., & Schurz, G. (2015). Qualitative probabilistic inference with default inheritance for exceptional subclasses. In C. Beierle, G. Kern-Isberner, M. Ragni, & F. Stolzenburg (Eds.), *Proceedings of the 5th Workshop on Dynamics of Knowledge and Belief (DKB-2015) and the 4th Workshop KI & Kognition (KIK-2015) co-located with 38th German Conference on Artificial Intelligence (KI-2015), CEUR Workshop Proceedings* (Vol. 1444).

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