Title
Can Formal Non-monotonic Systems Properly Describe Human Reasoning?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Kuhnmuench, Gregory
Ragni, Marco

Publication Date
2014

Peer reviewed

Can Formal Non-monotonic Systems Properly Describe Human Reasoning?

Gregory Kuhnmünch (gregory@cognition.uni.freiburg.de)
Department of cognitive Science, Friedrichstraße 50, 79098 Freiburg, Germany

Marco Ragni (ragni@cognition.uni.freiburg.de)
Department of cognitive Science, Friedrichstraße 50, 79098 Freiburg, Germany

Abstract

Monotonic (logical) reasoning makes the strong claim that an inference cannot be contradicted by future information; an assumption contrary to everyday life experience. This assumption is relaxed in non-monotonic reasoning. However, there are only few formal non-monotonic theories of reasoning that have inspired psychological theory-building. Can formal systems such as cumulative logic (system C) or preferential logic (system P), developed in philosophy and artificial intelligence, predict human non-monotonic inferences? Previous investigations have mainly used probabilistic representations of these systems and it remains unclear whether participants make the same inferences based on a qualitative description. We describe a different methodological approach and report related experimental findings that run counter to current approaches to human non-monotonic reasoning. Implications of our proposed method are discussed.

Keywords: human rationality; non-monotonic reasoning; belief revision; decision experiment; systems C, CL, and P

Introduction

Non-monotonic reasoning (NMR) is important for artificial intelligence (AI), but indispensable for everyday human reasoning. When we derive new information, we are often aware of the fact that information acquired later on can contradict previous conclusions or existing knowledge. Therefore, we are forced to resolve the contradiction to an extent that allows us to act efficiently in the world. This holds for rules of deontic reasoning, too: In what circumstances am I allowed to cross a red traffic light? In daily life we have to deal with exceptions from otherwise predominantly valid rules (Do not cross when the traffic light is red). Other domains are naïve psychology and theory of mind: Our initial assumptions about another's thoughts, emotions, or motives require revision once we have made new inconsistent observations of that person's behavior. Hence, everyday thinking is often non-monotonic (Pollock, 2008, only abstract available) – it requires NMR and dealing with exceptions. Similarly, expert systems or databases in AI might have to address such problems and therefore, classical designs are augmented by NMR-capabilities. Let us define a logic as non-monotonic if the set of (logical) conclusions from a theory (or knowledge base) is not necessarily preserved when new information is added to the theory. Previous conclusions or premises (declarative knowledge) might be retracted, similarly to belief revision (cp. Kraus, Lehmann & Magidor, 1990). Retraction means that their validity is lost – they are removed because their correctness is not warranted any more.

As a central result of psychological research take Byrne's (1989) suppression task: New knowledge about alternatively sufficient or additionally necessary premises can modulate validity of conclusions w.r.t. propositional logic. How could a theory describe human NMR? There are formal (non)-monotonic systems from AI (e.g., Kraus et al., 1990) and psychology (e.g., Pfeifer & Kleiter, 2005) that describe which conclusions can be derived under differing rationality assumptions. In AI, these assumptions are derived from logic; in psychology the standard is typically laymen's performance. Humans deviate from propositional logic (e.g., in the suppression task) and, more generally, conditional reasoning. Nonetheless, there are many other logics and there is a claim according to which non-montonic logics can (substantially) describe these findings (Dietz, Hölldobler, & Ragni, 2012; Stenning & van Lambalgen, 2006; 2008). Many other proponents investigate NMR systems: Benferhat, Bonnefon and Da Silva Neves, 2005; Elio and Pelletier (1997); Ford (2004), only to name a few.

Table 1: Rules of propositional logic and of systems C, CL, and P. OR and D are only valid in system P; LP is only valid in CL. Refl is reflexivity, SupCl is supraclassicality.

System	Rules
Propositional Logic	**P** + MT, CP, TT, EHD
C (**C**umulative)	Refl, LLE, RW, CT, CM; EV; AND, MPC, SupCl
CL (**C**umul.+**L**oop)	**C** + Loop
P (Preferential**)**	**C** + OR, S (= HHD), D (proof by case)
Extensions of P	DR, RM etc.

As we cannot describe the properties of all existing systems, we will focus on (i) three systems relevant for our experiment, that is systems C, CL, and P and (ii) the following rules: Loop (LP), Left Logical Equivalence (LLE), Right Weakening (RW), Cut (CT), Equivalence (EV), AND (AND), Modus Ponens in the Conclusion (MPC), Contraposition (CP), Transitivity (TT), OR (OR), Proof by Case (D), Disjunctive Rationality (DR), Rational

Monotonicity (RM), Monotonicity (MT), Cautious Monotonicity (CM), Easy half of deduction (EHD), and S (S).Table 1 shows which rules are comprised by systems C, CL, and P, and central rules of propositional logic.

Please note that monotonicity (MT) and transitivity (TT) are valid in propositional logic, but not in C, CL or P. Due to monotonicity, propositional logic does not allow for NMR. Read $\alpha \mid\!\sim \beta$ as "If α then normally β;" $\alpha \rightarrow \beta$ and $\alpha \models \beta$ as "If α then β." The symbol "∨" is inclusive logical "or" and "∧" is logical "and". Please note that $\alpha \models \beta$ implies $\alpha \mid\!\sim \beta$ but not vice versa. Classically, inferences are presented by an inference scheme, consider the monotonicity inference:

Monotonicity (MT)

$$\frac{\models \alpha \rightarrow \beta, \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma}$$

Read the formula above as "If α then β" and "if β then normally γ ," then conclude "if α then normally γ ." Other relevant rules are TT, CM and OR:

Transitivity (TT)

$$\frac{\alpha \mid\!\sim \beta, \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma}$$

Cautious Monotonicity (CM)

$$\frac{\alpha \mid\!\sim \beta, \alpha \mid\!\sim \gamma}{\alpha \wedge \beta \mid\!\sim \gamma}$$

OR

$$\frac{\alpha \mid\!\sim \gamma, \;\beta \mid\!\sim \gamma}{\alpha \vee \beta \mid\!\sim \gamma}$$

The remaining formulae and the complete experimental tasks can be found on the website[1]. System P (e.g., Kraus et al., 1990) contains "inference rules that are widely accepted as being the minimal set for any "reasonable" non-monotonic reasoning system" (Ford, 2004, p. 94).Pfeifer and Kleiter (2005) investigated three rules of P and their findings support system P. To clarify our methodological approach to NMR, some key aspects should be addressed: We propose (a) investigating a comprehensive set of the rules that are warranted (valid) within each of the systems, and (b) including rules of monotonic logic (as did Pfeifer & Kleiter, 2006; e.g. MT and TT). Such an approach yields a more comprehensive picture of human (non-)monotonic reasoning and it shows more precisely how NMR deviates from propositional logic.

Research on content effects shows the importance of declarative knowledge in many fields of human reasoning (see Beller & Spada, 2003), e.g., in the causal domain (Kuhnmünch & Beller, 2005). If content-related knowledge supports conclusions intended by experimenters, its effects are facilitating, otherwise they are rather inhibitory. With facilitating content effects, conclusions can be derived by mere recall. In this case, there is no need for a deliberate reasoning process. The question is whether NMR can be demonstrated with naïve reasoners beyond content effects. Do reasoners possess this abstract competency when contents and knowledge are minimal? Abstracted scenarios are well-suited to this end. For instance, Benferhat et al. (2005) described new life forms in the Arctic Ocean. In such scenarios only general biological knowledge can be brought to bear, nothing specific is said regarding the life form. Such abstracted scenarios imply minimal knowledge, but as opposed to completely abstract material, they still use concrete (and thus imaginable) propositions. To illustrate the distinction: *If (a is x) then (a is y)* is abstract, whereas *If the figure is red then the figure is square* is an abstracted version.

NMR is psychologistic – a measure of correctness does not depend on formal standards, but on laymen's performance (see Pelletier and Elio, 2005). As a consequence, empirical data from ordinary people is the very data that a formal non-monotonic logic must cover. Formal standards do not, however, become obsolete – they can serve as a frame of reference to determine differences between formal and human NMR. Of course, formal theories are rather an inspiration for a psychological theory. For example, they do not take restrictions of working memory into account, but a cognitive model can (cp. Stenning & van Lambalgen, 2008).

Regarding our hypotheses, the first refers to the new materials we used: Keywords like *normally* trigger non-monotonic reasoning patterns with abstracted tasks. That is, conclusions licensed by unwarranted rules such as MT and EHD should be drawn less frequently than those of rules warranted by the three systems. Furthermore, in line with Pfeifer and Kleiter (2005), we expect system P to yield better results than C and CL (frequency of rule use). Second, the transitivity rule (TT) is so familiar that it is expected to be applied nearly automatically by most participants. Finally, we expect difficulties to understand the scope of negation with the term *normally*: Statements of the form *A is not normally green* and *A normally is not green* are considered equivalent. The reason for this expectation is that in everyday life, *not normally green* is rarely used and thus it might be confused with *normally not green* (at least this holds for the German formulations *nicht normalerweise grün* and *normalerweise nicht grün*).

Method

Materials

Inference Tasks Instructions and formulation of tasks were optimized in two pre-tests. The main finding was that a conditional task format (premises like *If the figure is circular, then it is green*) was harder to comprehend. Hence, we removed the conditionals for the main test and used an equivalent relational description.

In addition to the rules mentioned in the introduction, some equivalent reformulations of rules were used to see whether the form of premises makes a difference: Modus

[1] http://webexperiment.iig.uni-freiburg.de/system.zip

Ponens in the Conclusion (MPC'), Contraposition (CP'), Disjunctive Rationality (DR'), Monotonicity (MT'), Rational Monotonicity (RM', RM''), and Easy Half of Deduction (EHD'). Neither supraclassicality (SupCl) nor reflexivity (Refl) were tested, as they belong to each of the systems C, P, and CL. Each task comprised a set of premises and two main answer options: (i) Yes, there is a conclusion that applies with certainty and (ii) No, nothing can be concluded with certainty. If participants ticked yes, then they had to mark the conclusion in the "building set" printed directly below. The set consisted of all figures and properties that appeared in the premises. A translated sample task for rule Monotonicity (MT) looks like this:

<u>Premises:</u>
The green figures are normally checkered.
The circular figures are normally green.

Is there a conclusion that applies with certainty if you assume that the premises are true?

□ **Yes**, this conclusion *applies* with certainty (check exactly one box per column):

□ The circular □ The green □ The checkered	figures are	□ in each case □ normally	□ circular □ green □ checkered

□ **No**, nothing can be concluded with certainty.

Instructions for Inference Tasks First, the main answer options and use of the building set were explained to participants; if they chose *yes* as a main answer option, they had to tick it. In each row of tick boxes within the building set, exactly one tick was required. Otherwise they had to tick the other main answer option *no*. A concrete sample that dealt with health conditions and mental fitness followed. Participants were instructed not to repeat any premise as answer and to provide the most informative conclusion. The latter requirement was to ensure that among alternative conclusions warranted by a system, the one with the most new information was marked in the building set. Finally, participants were informed that the following tasks combine certain properties of figures:

Shape: square, circular, triangular,…
Color: violet, green, blue,...
Filling: shaded, checkered, filled in...

Any combination of shape, color, and filling is possible, as well as further instances of these properties (ellipses). Therefore, the instruction did not establish a closed world with a restricted set of properties. For example, with a closed world limited to two shapes, the absence of one shape would justify inferring the other. To foster participants' imagination with these rather abstract tasks, they were given a fictional scenario: "Imagine you are a visitor at a factory observing the properties of produced artifacts."

Additional Task A second type of task examined the perceived semantics of negations in four rules:

(1) The circular figures are *not* normally green.
(2) The circular figures normally are *not* green.
(3) The circular figures are *not* in each case green.
(4) The circular figures are, in each case, *not* green.

Participants were asked *Which of these statements are equivalent?* In a table-like format, they had to tick equivalent pairs or omit ticks whenever they supposed there was no equivalence. By making no ticks, they could indicate that they thought there is no equivalent pair at all.

Design

We used a within-participant design in order to assess individual answering patterns across tasks. Participants received questionnaires that were the same except for the ordering of tasks: We used nine orders of tasks for balancing. Task orders were generated randomly, except that groups of very similar tasks did not appear on consecutive pages of the questionnaire, but were always separated by at least one task not belonging to that group. This is due to the fact that premises of these tasks were similar and in pretesting, some participants erroneously thought these tasks were identical. This led some of them to simply repeat the former inference or to omit the task. The groups of similar tasks were: (a) LP, RW, MT, MT', TT (b) LP, LLE, EV (c) MPC, MPC', EHD, EHD', S. Each order ended with tasks RM''-RM, as pretesting had shown very low frequencies of warranted answers with these tasks (RM: 16.67%, RM'': 0%). Hence, these tasks seemed particularly problematic for participants and we mused that they might influence consecutive tasks to a higher degree than others. In order to avoid monotony and carry-over effects, half of the tasks comprised the object properties square-violet-shaded and the other half circular-green-checkered. The additional task always came last for all participants.

Participants

Participants were informed about the experiment by flyers on the campus of the University of Freiburg. Altogether, 31 persons filled out the questionnaire, five of them were excluded as they did not fulfill the requirements: They had participated in the pretest and only informed experimenters after debriefing. Some had expert knowledge in propositional logic. The remaining 26 participants fulfilled the sampling requirements, among them 21 women and five men. They were all native German speakers. Mean age was

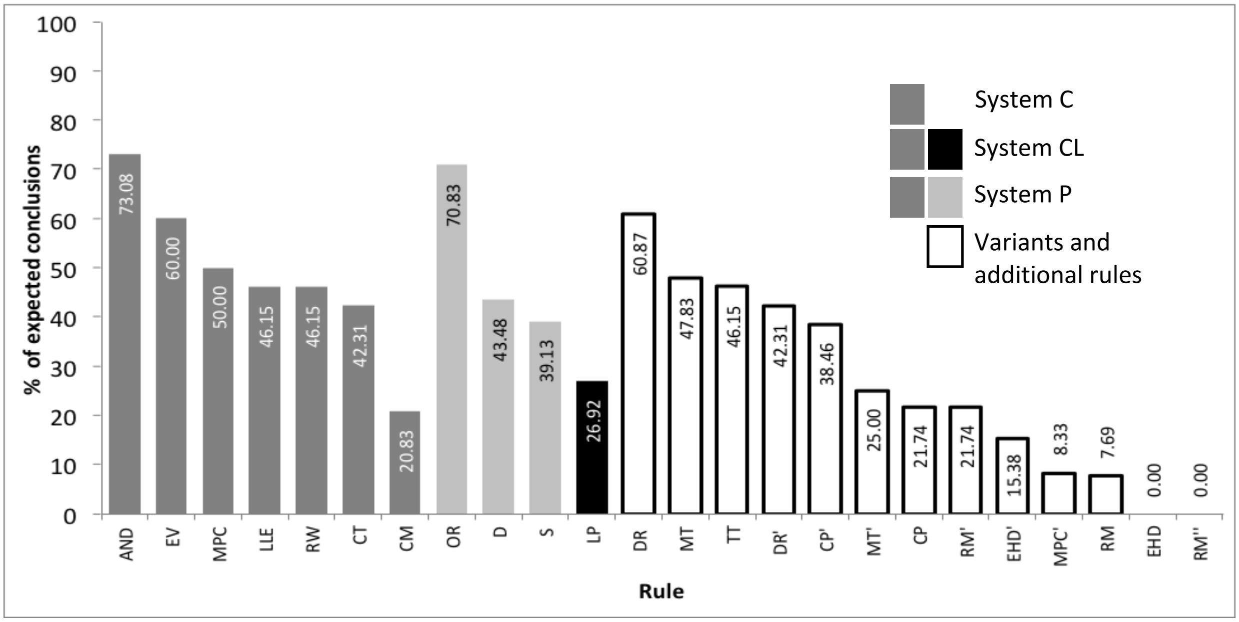

Figure 1: Relative frequencies of warranted conclusions by rule and formal systems by rules (see legend)

24.31 years (*SD* = 5.57). All of them were students of the University of Freiburg (Germany). Their average number of college semesters was 5.36 (*SD* = 5.71).

Procedure

Participants filled out the printed questionnaire in a quiet room in small groups of no more than five persons. First they worked on the inference tasks; then on the additional task. They could ask the experimenter questions if an instruction was unclear. Materials were presented in German. On average, they took about 45 minutes to fill out the questionnaire (excluding a short briefing and debriefing). For compensation, they received course credits or seven Euros.

Results

Order Effects

A logit-loglinear analysis was conducted with the variable task positions within the questionnaire as a factor and the dichotomous dependent variable warranted vs. unwarranted conclusion according to the rule. The factor could be removed from the model without significant loss of fit ($\chi^2(21) = 19.446$, $p = 0.557$).

Frequencies of Conclusions

Altogether, out of 26 (participants) * 24 (rules per questionnaire) = 624 answers, 614 were included into analyses (with ten answers missing or ambiguous). For the included answers, we calculated the proportion of warranted conclusions per rule (w.r.t. the systems of formal reasoning). Figure 1 shows the results. Using binomial tests we determined whether there was a significant difference between warranted and unwarranted conclusions for each rule. Only the AND-rule yielded such a result in favor of the conclusions (testing against an equal distribution, two-tailed, $p = .029$). CM, Loop and MPC' on the other hand were significantly more often divergent from the warranted conclusions ($n = 24$, $p = .007$; $n = 24$, $p = .029$; $n = 24$, $p = .001$, respectively).The remaining rules failed to reach significance in either direction ($p > 0.064$ in each case), half of the reasoners drew the warranted conclusion, the others did not. Notably, the OR-rule with 70.83% was closest to the result of AND (73.08%). On average, 58.45% of the answers were in accord with systems C, CL, and P; the remaining answers were divergent from the rules' formal predictions.

Comparison of Systems C, CL, and P

Systems can be compared by adding up frequencies for all conclusions warranted by a system (cp. the introduction): Conclusions in accord with a system score one; one is also scored when participants avoided a conclusion that is in disaccord with a system. All other answers yielded a zero scoring and did not contribute to the overall sum. The sums were 354 (C), 342 (CL), 349 (P). A χ^2-Test yielded no significant differences between the three systems ($\chi^2(2) = 0.209$, $p > .05$). The result is unchanged when the sums are normalized by the number of rules in support of each single system. Notably, the maximal number of rules answered in accord with each single system was 18 (C: two participants, CL and P by the same participant). A second way to look at the results is to focus on the three rules that differentiate the systems, namely D, LP and OR. Binomial tests showed that a significant proportion of answers diverged from the LP rule, $n = 26$, $p < .05$ (two-tailed binomial test). As reported before, the D- and OR-rules did not reach significance.

Transitivity Fallacy

The TT rule (unwarranted by any NMR-system) was used by 46.15% of participants. A binomial test did not reach significance, $n = 26$, $p > .05$. Thus, about half of the participants used this rule whereas the others refrained from any conclusion or rarely concluded something that deviated from the rule´s conclusion.

Shift of Scope with Negation

We turn to the four rules of the additional task (see materials section). Rules 1 and 2 tested whether participants differentiate between *not normally x* and *normally not x*. Twenty of 26 participants were in favor of an equivalence between rules 1 and 2, $p < .01$, two-tailed binomial test. We also used two-tailed binomial tests for the remaining comparisons with equal distributions: The corresponding variation in negation using *in each case* instead of *normally* yielded the opposite result: Twenty-five out of 26 answers denied an equivalence of *not in each case x* (rule 3) and *in each case not x* (rule 4), $p < 0.01$. Most participants (16 out of 26) stated that *not normally x* is equivalent to *not in each case x*, but the difference between these rules 1 and 3 was not significant, $p > .05$. Similarly, 15 out of 26 answers state an equivalence between *normally not x* and *not in each case x* (rules 2 and 3, $p > .05$). Finally, 24 of 26 participants denied an equivalence between *normally not x (rule 2)* and *in each case not x (rule 4)*, a significant result, $p < .001$.

Discussion

The challenge of the reported experiment was to activate the non-monotonic mode of reasoning within our participants using only keywords like *normally*. Our materials did so: An average of 58.45% of answers was in accord with the three systems. Nonetheless, some rules of propositional logic seem to be too familiar to not be used. Accordingly and in line with expectations, the transitivity (TT) rule was used in 46.15% of cases although it is not compatible with the three tested non-monotonic systems. The use of such unwarranted rules might have happened for an additional reason: We used abstracted scenarios (geometric figures, their fillings and colors in a factory's production), that is exceptions might not have been salient enough in all tasks. For a comprehensive study of rule systems, this was the method of choice, though: We were not interested in content effects that facilitate or inhibit reasoning with exceptions; rather, the research question focused on whether NMR exists as a competence beyond content effects — in a similar fashion as naïve reasoners can apply modus ponens and to a lesser degree modus tollens with abstract conditionals, e.g. in the Wason selection task. Our participants could certainly have imagined more concrete scenarios in order to solve the tasks, but with up to three premises per rule and the variety of figure properties mentioned therein, this is not easy. In follow-up experiments we intend to integrate the question of whether content effects trigger different modes of NMR. Content-related and rather (domain-)general aspects should be disentangled experimentally, otherwise competencies related to contents and to reasoning become mixed up (Beller & Spada, 2003; Kuhnmünch & Beller, 2005).

Two main results show the overall lack of support for any of the formal systems tested. Comparison of the whole set of rules (24) bore no advantage for any of the systems C, CL or P. On the other hand, there is much overlap of C and CL (only rule LP differentiating between them) and both overlap largely with P. As an alternative, the three rules D, LP and OR can be used to differentiate between the systems. As LP was rarely applied in the experiment, systems C and P remain the ones with the best fit for about two thirds of the answers (cp. Figure 1).

Certainly, attempts to corroborate some of the formal systems always need to focus on the subset of tasks that allows for a differentiation of accounts. This can tempt researchers not to examine whether participants conform with other aspects (rules) of a system or with propositional logic. However, we examined a more comprehensive set of rules, including rules not compatible with any of the systems (e.g., monotonicity).With fewer rules, discrepancies from a system's norm are less likely to be discovered. Although system P so far has more empirical support than C or CL, our results demonstrate that it does not surpass system C. Only system CL seems inappropriate as it postulates the LP-rule that was used rarely (26.92%).

The substantial proportion of unwarranted rules used for the conclusions also shows that human reasoning is indeed flexible and manifold – the same participants gave non-monotonic and monotonic answers. We agree with Stenning and van Lambalgen (2005) that rather the whole process of task interpretation and reasoning must be considered. Interpretation determines whether a non-monotonic reasoning process is activated or not. In the light of our results, descriptively more suitable and more flexible systems beyond pure formal systems are needed to describe the bandwidth of human NMR. As system P is sometimes considered the minimal set of rules that can describe human NMR and our participants applied only some of the rules, system O is a candidate. It comprises fewer rules and might be augmented by additional rules independently of the rather rigid set of rules adopted by formal systems. This liberal attitude towards rule systems might not be licensed by formal logics, but is absolutely permissible if the objective is a model of human reasoning. Then, even human inconsistencies should be modeled.

The additional task demonstrated a strong ambivalence of negations in materials framed in natural language. Therefore, it should be considered whether a disambiguation is possible by modifying materials or by training participants regarding the intended meaning. Another view on these results is closed world reasoning (e.g., Stenning & van Lambalgen, 2008, p. 36ff), in the sense of regarding the properties named in the task as the complete set of possibilities, that is all figures have either the one property or the other, but not a third.

Our questionnaire discouraged participants from adopting such a view, though: Remember that the instruction named three geometric forms, three colors, and three fillings for objects. Furthermore, ellipses after these properties made clear that there might be more instances. The rules (tasks) used only two properties of each triplet; hence, the instruction did not suggest a closed world. Nevertheless, some participants could have adopted this interpretation, which would justify equivalences such as *not green* means *violet*. Rules like

(1) The circular figures are *not* normally green.

(2) The circular figures normally are *not* green.

might then be interpreted and transformed first: *normally are not green* (2) means *normally are violet*. And *not normally green* (1) means *rarely green* which yields *normally violet* by the assumption that the violet and green objects sum up to 100% of the geometric figures (another way to apply the closed world). As a result, we get:

(1) The circular figures normally are violet.

(2) The circular figures normally are violet.

From this perspective, the scope shift with negation might be justified and not be seen as a failure to correctly interpret the scope of negation. We believe these conflicting explanations deserve further investigation due to their importance for NMR-tasks.

An equivalence of *normally* (rules 1 & 2) and *in each case* (rules 3 and 4) cannot be explained by a closed world, though. There seems to be a different semantic understanding: Either both terms are taken by participants to mean exactly 50% of the cases or they are both taken to mean the same numerical majority of the cases (e.g., 80%).

The methodology we used – a rich set of rules and analysis beyond a few differentiating rules – is suggested as a complementary paradigm to shed more light on the appropriateness of formal systems as approximations of human NMR. All accounts should be enriched by systematic empirical research regarding content effects and closed world reasoning and task interpretation in a broader sense, and most importantly to let reasoners generate conclusions instead of simply verifying given conclusions. Disentangling these aspects is essential for the development of valid psychological accounts of human NMR – probably as essential as it has proven for human monotonic reasoning.

Acknowledgments

We would like to thank Laura Schumacher, Julia Grzimek, Ivanessa Mirastschijski, and Mathis Tometten for their valuable help with data collection and raw data extraction. Mathis Tometten and Antonia Becher helped with data analysis and the manuscript. In particular, we would like to thank Dr. Stefan Wölfl for his invaluable help in generating hypotheses, formal aspects, experimental materials, and in-depth discussions. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program *New Frameworks of Rationality* (SPP 1516, project *Models of human non-monotonic reasoning*).

References

Beller, S., & Spada, H. (2003). The logic of content effects in prepositional reasoning: The case of conditional reasoning with a point of view. *Thinking & Reasoning*, *9*(4), 335-378.

Benferhat, S., Bonnefon, J. F., & da Silva Neves, R. (2005). An overview of possibilistic handling of default reasoning, with experimental studies. *Synthese*, *146*(1-2), 53-70.

Byrne, R. M. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*(1), 61-83.

Dietz, E. A., Hölldobler, S., & Ragni, M. (2012). A Computational Approach to the Suppression Task. In *Proceedings of the 34th Cognitive Science Conference*.

Elio, R., & Pelletier, F. (1997). Belief change as propositional update. *Cognitive Science*, *21*(4), 419-460.

Ford, M. (2004). System LS: A Three-Tiered Nonmonotonic Reasoning System. *Computational Intelligence*, *20*(1), 89-108.

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

Kuhnmünch, G., & Beller, S. (2005). Distinguishing Between Causes and Enabling Conditions: Through Mental Models or Linguistic Cues? *Cognitive Science*, *29*(6), 1077-1090.

Pelletier, F. J., & Elio, R. (2005). The case for psychologism in default and inheritance reasoning. *Synthese*, *146*(1-2), 7-35.

Pfeifer, N., & Kleiter, G. D. (2005). Towards a mental probability logic. *Psychologica Belgica*, *45*(1), 71-99.

Pfeifer, N. & Kleiter, G. D. (2006). Is human reasoning about nonmonotonic conditionals probabilistically coherent? In *Proceedings of the 7th Workshop on Uncertainty Processing* (p. 138–150). Mikulov (Czech Republic).

Pollock, J. L. (2008). Defeasible reasoning. In J. E. Adler, L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations*. New York, NY US: Cambridge University Press.

Stenning, K., & van Lambalgen, M. (2005). Semantic Interpretation as Computation in Nonmonotonic Logic: The Real Meaning of the Suppression Task. *Cognitive Science*, *29*(6), 919-960.

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