Proceedings of ICCM 2020

18th International Conference on Cognitive Modelling[1]

Edited by

Terrence C. Stewart

1 Co-located with the 53rd Annual Meeting of the Society for Mathematical Psychology and held online

Preface

The International Conference on Cognitive Modeling (ICCM) is the premier conference for research on computational models and computation-based theories of human cognition. ICCM is a forum for presenting and discussing the complete spectrum of cognitive modelling approaches, including connectionism, symbolic modeling, dynamical systems, Bayesian modeling, and cognitive architectures. Research topics can range from low-level perception to high-level reasoning. In 2020, ICCM was jointly held with MathPsych – the annual meeting of the Society for Mathematical Psychology. While ICCM and MathPsych were originally to be held in Toronto, Canada from July 25th-28th, the ongoing COVID-19 pandemic made this impossible. Instead, the conference was held online from July 20th to July 31st, using a combination of pre-recorded videos, live discussions, and custom software developed by the Society for Mathematical Psychology.

Acknowledgements

We would like to thank the Society for Mathematical Psychology (SMP) for their tremendous work in converting this conference to an online venue. In particular, the MathPsych Conference Chair (Joachim Vandekerckhove) guided the development of a custom website, and the officers of the SMP (especially Leslie Blaha) gave much-needed logistical support. ConfTool was used to manage submissions and reviews.

Papers in this volume may be cited as:

Lastname, A., Lastname, B., & Lastname, C. (2020). Title of the paper. In Stewart, T. C. (Ed.). *Proceedings of the 18th International Conference on Cognitive Modelling* (pp. 6-12). University Park, PA: Applied Cognitive Science Lab, Penn State.

ISBN-13: 978-0-9985082-4-5, published by the Applied Cognitive Science Lab, Penn State.

(C) Copyright 2020 retained by the authors

The Power of Nonmonotonic Logics to Predict the Individual Reasoner

Marco Ragni (ragni@cs.uni-freiburg.de)
Cognitive Computation Lab, University of Freiburg, Germany

Francine Wagner (wagner.francine94@gmail.com)
Cognitive Computation Lab, University of Freiburg, Germany

Sara Todorovikj (sara.todorovik@gmail.com)
Cognitive Computation Lab, University of Freiburg, Germany

Abstract

Human reasoning systematically deviates from conclusions predicted by classical logic. It is *nonmonotonic* and *defeasible*, i.e., new information can lead to the retraction of previous inferences. While these results hold for the analysis of population data, it is an open question, if nonmonotonic logics can capture individual human reasoning. In this article, we take three prominent nonmonontontic approaches, the Weak Completion Semantics, Reiters Default Logic, and OCF, a ranking on possible worlds, implement variants of them and evaluate them within the CCOBRA-framework for their predictive capability in the Suppression Task. We demonstrate that nonmonotonic approaches are able to predict individual reasoner on 82% (median). Furthermore, we can demonstrate that OCF and an improved version of Reiter make identical predictions and that abduction is relevant on the level of an individual reasoner. We discuss implications of logical systems for human reasoning.

Keywords: Human Reasoning; Nonmonotonic Logic; Evaluation; Individual Reasoner

Introduction

Humans draw different conclusions when they are presented with additional information. Consider the following information:

> If Lisa has an essay to write then she will study late in the library.
> She has an essay to write.

In a study about 96% of participants concluded, that *she will study late in the library* (Byrne, 1989). However, only 38% of the participants receiving in the same study the additional conditional

> If the library stays open, she will study late in the library.

endorsed the conclusion that *She will study late in the library*. In contrast to about 90% of the reasoners endorsing the respective conclusion when receiving the alternative conditional

> If she has a textbook to read, she will study late in the library.

While this change in the inference behavior between the first case to the second case might be intuitively clear for the reader, because the second conditional provides a possible constraint, the inference *she will study late in the library* even with such an additional constraint is classically logically valid. Hence, this and other findings (e.g., in the Wason Selection Task, see Ragni, Kola, & Johnson-Laird, 2018) demonstrate that the normative framework of classical logic is not a good descriptive framework for about *how* humans reason. Despite that logical inference problems require that humans derive the logical conclusion. In recent years this has lead to a number of different modeling approaches, e.g., probabilistic models (Oaksford & Chater, 2013), heuristic models (Evans & Over, 2004), and recently to the discovery that nonmonotonic logics might be appropriate (Stenning & Lambalgen, 2008). A reasoning theory is called nonmonotonic if new information can lead to the retraction of previous inferences (Antoniou, 1997). Hence, the reasoning process allows for defeasible conclusions or reasoning under exceptions. In fact, many cognitive theories are nonmonotonic, e.g., probabilistic (Oaksford & Chater, 2013), mental models (Johnson-Laird, Girotto, & Legrenzi, 2004), and cognitive logics (Stenning & Lambalgen, 2008). In this article, we investigate three noteworthy logical theories are to predict an individual reasoner conclusion, before the reasoner generates it. The paper is structured as follows: First, we will introduce the necessary background for conditional reasoning and the Suppression Task. In the next section we introduce three prominent models of nonmonotonic reasoning. Then, a section with the evaluation framework CCOBRA and the description of a data-set follows. Finally, a section on the evaluation of the different nonmonotonic logics and their variants and a discussion about the implications concludes the paper.

Background & Related Work

Conditional Reasoning

Conditional reasoning (i.e., reasoning about "if") is diverse from reasoning about given facts. It can represent assumptions about states, e.g., about causal relations or in action planning, considering hypothetically potentially different past states (e.g., counterfactual reasoning), or hypothesizing theories (e.g., inductive reasoning). It is highly relevant for both automated and human reasoning. There is a long history in cognitive science about modeling conditional reasoning, i.e., a statement of the form, "if e then l", often written by $e \rightarrow l$ or $(l|e)$. For a given conditional four inference mechanisms are possible:

Modus ponens (MP) is a deductively valid argument in classical logic: from the premises $e \rightarrow l$ and e, the consequent l is inferred. Consider the premises:

> If she has an essay to write then she will study late in the library. She has an essay to write.

The valid conclusion is:

> She will study late in the library.

Modus tollens (MT) is a deductively valid argument in classical logic: from the premises $e \rightarrow l$ and $\neg l$ the negated antecedent $\neg e$ is inferred. An example:

> If she has an essay to write then she will study late in the library. She will not study late in the library.

The valid conclusion is:

> She does not have an essay to write.

Denial of the antecedent (DA) is a deductively invalid argument in classical logic: from the premises $e \rightarrow l$ and $\neg e$ the negated consequent $\neg l$ is inferred. An example:

> If she has an essay to write then she will study late in the library. She does not have an essay to write.

The erroneous conclusion is:

> She will not study late in the library.

Affirmation of the consequent (AC) is also a deductively invalid argument in classical logic: from the premises $e \rightarrow l$ and l, the antecedent e is inferred. An example:

> If she has an essay to write then she will study late in the library. She will study late in the library.

The erroneous conclusion is:

> She has an essay to write.

All inference mechanisms are logically valid in case of a biconditional interpretation, i.e., if and only if she has an essay to write she will study late in the library. It has been claimed that the nonmonotonic System P satisfies minimal rationality criteria. And, that it is close to human reasoning (Pfeifer & Kleiter, 2005). A different study could not support the relevance of the nonmonotonic System P as a good descriptive theory to explains psychological findings (Kuhnmünch & Ragni, 2014). So we exclude this theory.

Suppression Task

Together with the aforementioned Wason Selection Task, the Suppression Task (the difference in resoning behavior with or without the additional conditional) is one core problem for reasoning theories (Byrne, 1989; Neth & Beller, 1999; Chan & Chua, 1994; Politzer, 2005). Accordingly, this task and human nonmonotonic reasoning has been modeled by many researchers (Dietz, Hölldobler, & Ragni, 2012a; Stenning & Lambalgen, 2008). Recent research (Ragni, Eichhorn, & Kern-Isberner, 2016) analyzed if nonmonotonic systems have the competence to grasp the specific nonmonotonicty of the Suppression Task without additional background information. It demonstrated that the inference mechanisms of all nonmonotonic logics despite the weak completion semantics required additional knowledge. This was, however, an evaluation on the aggregated level. So theories and models were competent to solve the problems with some requiring additional background knowledge. So far no analysis on predictive performance of the cognitive models or logics for the individual human reasoner on conditional reasoning in the Suppression Task. Before we can do so we present three main theories for nonmonotonic reasoning.

Models of Nonmonotonic Reasoning

The Weak Completion Semantics

One main criticism against classical two valued approaches is that in everyday life we typically have degrees of (un-) certainty. The traditional two symbols $\top$, $\bot$, are extended with U denoting *true*, *false*, and *unknown*, respectively. Stenning and Lambalgen (2008) have claimed that conditionals should be encoded by "licenses for inferences". For example, the conditional *if she has an essay to finish, she will study late in the library* or short ($l \leftarrow e$) should be encoded by the clause $l \quad e \wedge \overline{ab}_1$, where ab_1 is an *abnormality* predicate which expresses that l holds if e holds and nothing abnormal is known. The programs obtained for the two main examples of the Suppression Task are depicted in the first two columns of Table 1.

The abnormality predicates (e.g., ab_1) represent abnormal cases: For instance, ab_1 is true when *the library does not stay open* and ab_3 is true when *she does not have an essay to finish*. Weak completion is the process of substituting the conditional with a biconditional.

In the case of AC where the conclusion holds the propositional variable e is mapped to unknown. Hence, if we observe that 'she will study late in the library', then we cannot explain by the model that 'she has an essay to write' without abduction (Saldanha, Hölldobler, & Rocha, 2017). Abduction searches for the minimal explanation. Since e is the only undefined propositional letter in this context, the set of abducibles is $e \leftarrow \top$, $e \leftarrow \bot$. The above observation can be explained by selecting $e \leftarrow \top$ from the set of abducibles, weakly completing it to obtain $e \leftarrow \top$ and adding this equivalence to the logic program.

Reiter's Model for Default Reasoning

Reiter (1980) proposed a system for default reasoning. According to Reiter, conditionals of the form $e \rightarrow l$ are interpreted as default rules, i.e., they are true as long as no exception is known. This idea was inspired that the conditional "if an animal is a bird, then it can fly" is true as long as we know that this animal is not a penguin (or any other exception such as a dodo etc.). For reasons of simplicity we do not introduce the formalizations of a background theory. A default rule constructed from a conditional has:

Table 1: The WCS approach to the suppression task. ELT = the statements 'essay', 'late', and 'textbook to read'; ELO = the statements 'essay', 'late', 'open hold'; *ab* are abnormality predicates; wc is the weak completion; and lm_L are the least model; Table adapted from Dietz et al. (2012b).

Premise	$\mathcal{P}_{ELT}$	$\mathcal{P}_{ELO}$
Clauses	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow t \wedge \overline{ab}_2$ $ab_1 \leftarrow \bot$ $ab_2 \leftarrow \bot$ $e \leftarrow \top$	$l \leftarrow e \wedge \overline{ab}_1$ $l \leftarrow o \wedge \overline{ab}_3$ $ab_1 \leftarrow \overline{o}$ $ab_3 \leftarrow \overline{e}$ $e \leftarrow \top$
$\mathsf{wc}\,\mathcal{P}$	$l \leftrightarrow (e \wedge \overline{ab}_1)$ $\vee (t \wedge \overline{ab}_2)$ $ab_1 \leftrightarrow \bot$ $ab_2 \leftrightarrow \bot$ $e \leftrightarrow \top$	$l \leftrightarrow (e \wedge \overline{ab}_1)$ $\vee (o \wedge \overline{ab}_3)$ $ab_1 \leftrightarrow \overline{o}$ $ab_3 \leftrightarrow \overline{e}$ $e \leftrightarrow \top$
$\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}$	$\langle\{e,l\},\{ab_1,ab_2\}\rangle$	$\langle\{e\},\{ab_3\}\rangle$
$\mathsf{lm}_L\mathsf{wc}\,\mathcal{P}$	Infers l	Not infers l
Empirical Results Byrne (1989)	96 % infer l	38 % infer l

$$\frac{e : \textit{justifications}}{l}$$

- the precondition: e (an *e*ssay to write)
- the justifications: depends on the scenario, for the library scenario a justification is that the library is open.
- the consequence: l (study *l*ate in the library)

To model the Suppression Task we construct the default rules from the conditionals, i.e., for the conditional "if Lisa has an essay to write then she will study late in the library." we assume that no exception is known and so the exception above is empty. This changes when we learn about the exception that the library is not open, then the justification is that the library is open. Facts can be formulated as a conditional with a true antecedent."Lisa has an essay to finish"

$$\frac{:}{e}$$

This means that without any precondition and without any justification it can be inferred that Lisa has an essay to finish. Conditionals with tautology as antecedent are plausible statements or facts about the world (Beierle & Kutsch, 2019). Statements such as "Mostly, Lisa has an essay to finish" can be translated to:

$$\frac{: e}{e}$$

Table 2: The successive generation of ranks (ranks are represented by κ) of possible worlds for the conditionals 'if e then l' and 'if t then l'.

	e	l	t	$\kappa(\omega)$	$\omega_i \models (l\|e)$	$\kappa(\omega)$	$\omega_i \models (l\|t)$	$\kappa(\omega)$
ω_1	0	0	1	0	$\overline{e}$	1	$t\overline{l}$	3
ω_2	0	1	1	0	$\overline{e}$	1	tl	1
ω_3	1	0	1	0	$e\overline{l}$	2	$t\overline{l}$	4
ω_4	1	1	1	0	el	0	tl	0

the justification ensures that the cases where she does not have an essay to finish, are handled as an exception. Our implementation works as follows: After translating the fact and conditionals, the defaults are executed, starting with the fact and keeping the order of the conditionals from the original task. Since the defaults, constructed from the fact, do not have any precondition and our knowledge about the world (which is written as W in Reiter's terminology) is empty, they are always added to the world knowledge W.

Ordinal conditional function

A different approach is inspired by the relevance of worlds and so some worlds are more relevant than others. This inspired the idea that the relevance of the worlds impose a rank on the worlds. There are three kinds of worlds for a conditional $(l|e)$:

- worlds satisfying e and l, $\omega \models el$
- worlds falsifying the conditionals, assuming e is true but the consequence l to be false $\omega \models e\overline{l}$
- worlds assuming e to be false, $\overline{e}$, called inapplicable worlds.

An inapplicable conditional means that there can't be made any statement about l, as e is already assumed to be false. The different sets can be represented by an indicator function (Calabrese, 1991):

$$(l|e)(\omega) = \begin{cases} 1 & \textit{If } \omega \models el \\ 0 & \textit{If } \omega \models e\overline{l} \\ u & \textit{If } \omega \models \overline{e} \end{cases}$$

where u means undefined, i.e., a case where the precondition is false. This represents that when the precondition does not hold, a conclusion about the truth value of the consequence relation l cannot be made. The conditional $(l|e)$ is evaluated as true (has the value 1), if the possible world ω has e and l as true. In this case we write for this world el.

Instead of assigning probabilities to a world, we can use ordinal conditional functions (OCFs)

$$\kappa : \Omega \to \mathbb{N} \cup \{\infty\} \textit{ with } \kappa^{-1}(0) = \emptyset$$

which maps possible worlds to an integer value. They map any possible world $\omega \in \Omega$ to natural number, which represents the degree of disbelief (Spohn, 1988). They express degrees of plausibility of propositional formulas ϕ by specifying degrees of disbelief of their negation $\bar{\phi}$ (Beierle & Kutsch, 2019). The more plausible a world ω is, the smaller its rank $\kappa(\omega)$. Consequently $\kappa(\omega_1) = 0$ describes the world ω_1, which is the most plausible. Hence $\kappa(\omega_2) = 1$ is a world a little bit less plausible, than ω_2, whereas $\kappa(\omega) = 10$ or higher is a world very unplausible. At least one world needs to have the ranking of 0. $\kappa \models (l|e)$, *iff* $\kappa(el) < \kappa(e\bar{l})$ means the agent would conclude that the verification of the conditional is more plausible or less surprising than the falsification of the consequence (Beierle & Kutsch, 2019). The main idea of our model is that all possible worlds or assignments, are ranked by their plausibility. The most promising world, that matches our choices will be returned accordingly.

An example for computing ranks of worlds For every task, we assume that, in the beginning, all worlds are equally possible and have the rank 0 (cp. Table 2), since we do not know anything about our environment. The rank will be updated with multiple conditionals. Secondly all κ values for these sets are determined: $\kappa(el)$, $\kappa(e\bar{l})$, $\kappa(\bar{l})$. The κ values for the sets will have the same rank as the world with the smallest rank from each set accordingly. In Table 1 we have an example of possible worlds, which are revised with two conditionals. For the input:

> If Lisa has an essay to finish, she will study late in the library. She will study late in the library.

The possible worlds can describe an environment with 2 literals l and e. We write $\top$ for a tautology, i.e., the truth value *true*. The first sentence of the task is encoded to the conditional $(l|e)$ and the fact is encoded to a conditional without a precondition $(l|\top)$.

At the beginning all worlds are equally plausible, therefore $\kappa(\omega) = 0$. We use the first conditional $(l|e)$ to revise the belief and to update the κ values for each world. Therefore the worlds are split into three sets: verifying, falsifying and inapplicable worlds. Then, the variables $\kappa(el)$, $\kappa(e\bar{l})$, $\kappa(\bar{e})$ can be determined, all having a starting value 0. Finally, all worlds are updated accordingly and a new conditional can revise our belief in the same way the first one did. The second information is a fact and is transformed into a conditional with a precondition which is always true: $(l|\top)$. The values for $\kappa(el)$, $\kappa(e\bar{l})$, $\kappa(\bar{e})$ are computed. The most plausible world is the last one. When having the choice between:

> Lisa has an essay to finish
> Lisa has not an essay to finish

we will chose the first one, because it is consistent with the most plausible world: ω_4.

Individual Predictions

CCOBRA

The Cognitive Computation for Behavioral Reasoning Analysis (CCOBRA) framework is a benchmarking tool implemented in Python that actively integrates the individual human into the prediction loop. There is a close connection to psychological experiments. Implemented models are supposed to simulate the experimental procedure for individual participants. By providing responses to individual tasks, models are evaluated based on their predictive accuracies[1]. The CCOBRA framework offers multiple possibilities, e.g., a pretrain, adapt and predict methods that we used for our model evaluation.

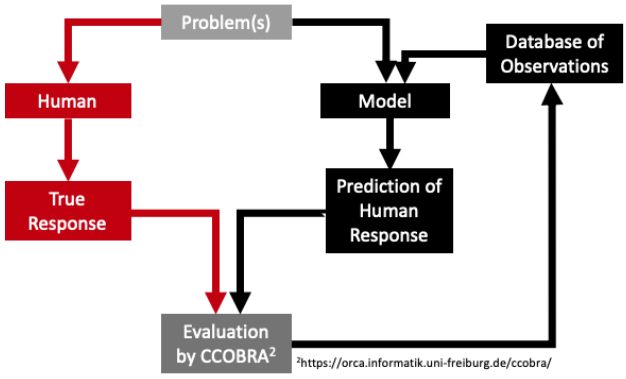

Figure 1: The CCOBRA-framework to evaluate the predictive power of cognitive theories.

Data-Set

The data can be found online[2]. It consists of 96 participants with no background in logic. Participants were recruited for a laboratory experiment at the University Freiburg. In 12 problems participants were requested to answer if specific conclusions follow from given information. Participants were divided into four groups. Group A received tasks with simple conditional arguments (non-suppression group). Group B also received tasks with simple conditional arguments but with a linguistic modification in premise one by adding the keyword "mostly" (they received the problem description in German).

> If Lisa has an essay to finish then she will mostly study late in the library. Lisa will study late in the library.
> Does Lisa have an essay to finish?

Group C received the modification in premise two by adding the keyword "mostly":

> If Lisa has an essay to finish then she will study late in the library. Mostly, Lisa will study late in the library.
> Does Lisa have an essay to finish?

[1] http://orca.informatik.uni-freiburg.de/ccobra

[2] https://github.com/CognitiveComputationLab/cogmods/tree/master/suppression_task

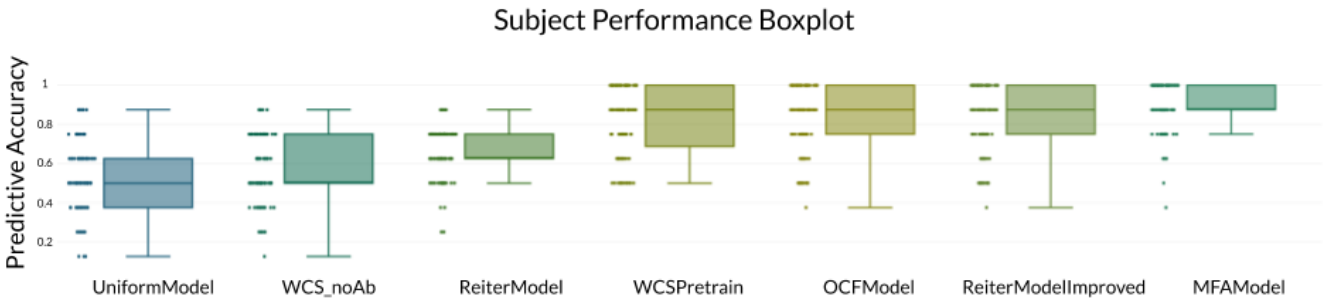

Figure 2: Boxplots for the models indicating individual subject performance. The data used for the plot are accuracies for each individual human reasoners performance (cp. Fig. 1.)

Group D received items with simple conditional arguments and two additional arguments:

> If Lisa has an essay to finish then she will study late in the library. If Lisa has some textbooks to read then she will study late in the library. If the library stays open then she will study late in the library. Lisa will study late in the library. Does Lisa have an essay to finish?

All four groups received three different scenarios (an abstract version, a story on an exoplanet with aliens having specific properties, and the library example above) and these in turn with the four inferential figures modus ponens (MP), modus tollens (MT), denial of the antecedent (DA) and affirmation of the consequent (AC). The arguments were presented sequentially and the concluding question had to be answered with "yes" or "no".

As our goal is to model the individual reasoner, we only report some aggregate statistic: Group A indicated the highest likeliness (88%) that the made conclusion holds true followed by Group D (80%), Group C (70%) and Group B (67%). For the logical correctness Group D shows the highest correctness rate (61.0%) followed by Group C (55.8%), Group A (54.9%) and then Group B (52.0%).

Predictive Performance of the Models

We compare now the predictive performance of the models with each other and baselines (cp. Fig. 2):

Baseline 1: Most Frequent Answer. To compare the cognitive models the most frequent answer is a good empirical measure. It represents the answers of all participants given for the same problem and returns the response which was answered the most. Taking the majority vote into account, it achieves a precision of 89.2%. This means that participants are quite homogeneous, i.e., they do not differ much in the responses they give.

Baseline 2: The Uniform Model. The *uniform random model* provides another base line, namely for participants that select randomly an answer. In the presented experiments, the participants had two choices for each task. So the uniform random model selects randomly one of the two choices and returns it as the predicted answer. It achieved, as expected, a 50% prediction performance.

The Weak Completion Semantics. This model which is based on a ternary logic and logic programs with allowing for abductive reasoning has a predictive performance of 82% based with an upper bound of 100% and a lower bound of 50%. If abductive reasoning is not allowed the performance drops to 56.1%, with a lower 12.5% and an upper bound of 87.5%.

The OCF-Model. The OCF model which is based on computed ranks of models reaches a high level of predictivity of about 82.2%. This model achieved an accuracy of 82.2%, with an upper bound of 100% and a lower bound of one single person with 37.5%.

The Classical Reiter Model. This is Reiter's original model Reiter (1980). It achieves a predictive accuracy of 65.5%, with predicting some persons as high as 87.5% and others as low as 25%.

Reiter Model Improved. The basic Reiter Model can be extended by adding default rules in order to model the phenomen that subject tend to use the modus tollens or affirmation of the consequent inference rule. By adding these two rules we reach the identical predictivity of the OCF-model with 82.2%, i.e., it predicts the exact same answer for every single subjects. This demonstrates a functional equivalence of the Reiter Model that is augmented with two additional rules with the semantic based approach by the OCF.

Discussion

Human reasoning has often been disqualified as "unlogical". While many psychological findings demonstrate that humans do deviate from valid inferences predicted by classical logic,

this paper demonstrates that nonmonotonic logics are competitive. They are even able to predict a median of 82% of the inferences drawn by every individual human reasoner in the Suppression Task. While the classical approach by Reiter did not match the high performance of the OCF model, we extended Reiter's model with two rules and demonstrated a functionally equivalent model to the OCF. The performance of the pre-trained version of the WCS model is only slightly lower than Reiter Model Improved and OCF Model. WCS deviated on some problems in the MP-case from the participants responses, where due to introduced abnormalities, it did not predict the classical MP conclusion. On the other hand, the WCS model successfully managed to model Denial of Antecedence problems by abductive reasoning in the cases of induced non-monotonicity, which the Reiter Model Improved and OCF did not succeed in. This analysis gives further support for the importance of abductive reasoning that has been reported relevant in the Weak Completion Semantics (Breu, Ind, Mertesdorf, & Ragni, 2019). Focusing on the individual predictivity of each system in a training set and a test set of participants in the CCOBRA-framework allows to estimate the true power of logics and cognitive models and makes even more progress possible, because it allows to identify individuals that are perfectly predicted and individuals in turn that are not captured. Future research needs to cover more experimental data, more cognitive theories, and aim to identify successful mechanisms of highly-predictive theories.

Acknowledgements

This paper was supported by a DFG-Heisenbergfellowship and DFG grants RA 1934/9-1, RA 1934/4-1, and RA 1934/3-1 to MR.

References

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

Beierle, C., & Kutsch, S. (2019). Systematic generation of conditional knowledge bases up to renaming and equivalence. In *European conference on logics in artificial intelligence* (pp. 279–286).

Breu, C., Ind, A., Mertesdorf, J., & Ragni, M. (2019). The weak completion semantics can model inferences of individual human reasoners. In *Logics in artificial intelligence - 16th european conference, JELIA 2019, rende, italy, may 7-11, 2019, proceedings* (pp. 498–508).

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

Calabrese, P. (1991). *Deduction and inference using conditional logic and probability* (Tech. Rep.). Naval Ocean Systems Center San Diego CA.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: syntactic views, mental models, and relative salience. *Cognition*, *53*(3), 217 - 238.

Dietz, E.-A., Hölldobler, S., & Ragni, M. (2012a). 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., & Ragni, M. (2012b). A Simple Model for the Wason Selection Task. In T. Barkowsky, M. Ragni, & F. Stolzenburg (Eds.), *Report Series of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition.*

Evans, J. S. B. T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, *111*(3), 640.

Kuhnmünch, G., & Ragni, M. (2014). Can Formal Nonmonotonic 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* (p. 1222-1228). Austin, TX: Cognitive Science Society.

Neth, H., & Beller, S. (1999). How knowledge interferes with reasoning-suppression effects by content and context. In *Proceedings of the twenty first annual conference of the cognitive science society* (pp. 468–473).

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, *19*(3-4), 346-379.

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 Homan Inferences in the Light of New Information: A Formal Analysis. In *Proceedings of the 25th international joint conference on artificial intelligence (ijcai'16).*

Ragni, M., Kola, I., & Johnson-Laird, P. N. (2018). On selecting evidence to test hypotheses: a theory of selection tasks. *Psychological Bulletin*, *144*(8), 779-796.

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

Saldanha, E.-A. D., Hölldobler, S., & Rocha, I. L. (2017). The weak completion semantics. In *Bridging@ cogsci* (pp. 18–30).

Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in decision, belief change, and statistics* (pp. 105–134). Springer.

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