Asking Human Reasoners to Judge Postulates of Belief Change for Plausibility

(Extended Abstract)

Clayton K. Baker*, Thomas Meyer

University of Cape Town and Centre for Artificial Intelligence (CAIR), Cape Town, South Africa

Abstract

Empirical methods have been used to test whether human reasoning conforms to models of reasoning in logic-based artificial intelligence. This work investigates through surveys whether postulates of belief revision and update are plausible with human reasoners. The results show that participants' reasoning tend to be consistent with the postulates of belief revision and belief update when judging the premises and conclusion of the postulate separately.

Keywords

revision postulates, update postulates, human reasoning, survey

1. Introduction

It has been shown that human reasoning displays non-monotonicity, but the methodologies that test this relationship differ within the AI community. An example [1] of one approach is through surveys in which English translations of the postulates of defeasible reasoning were judged for plausibility by human reasoners. As another example, a combined approach [2] was also used to investigate the link between formal theories of non-monotonic reasoning and the extent to which humans reason defeasibly. The combined approach involved a theoretical and empirical analysis. In the theoretical analysis, the predictions of each system was compared using the Suppression Task [3], a logical experiment used in the psychology community in which subjects appear to retract valid logical inferences when subjects gain new information. In the empirical analysis, three experiments were used to test the predictions of each system, as well as the inferences of human reasoners, with strict and defeasible knowledge. While there are empirical studies that investigated the relationship between non-monotonic reasoning with human reasoning, the relationship between belief change and human reasoning has been primarily studied from a theoretical perspective, e.g. in classical logic [4], probability and possibility theory [5], ontologies [6] and abstract argumentation [7]. Our first hypothesis is that human reasoning is consistent with postulates, advanced by Alchourrón, Gärdenfors and Makinson (AGM) [8], for belief revision. To enable comparison, we also hypothesise that human reasoning is consistent with postulates, advanced by Katsuno and Mendelzon (KM) [9], for reasoning with belief update. We investigate the hypotheses at the postulate level and at the system level. We use the language of propositional logic in the formulation of our postulates to construct logically closed belief sets. Additionally, we note that once our hypotheses are tested using propositional logic as the underlying language, our results can be lifted to other forms of logic. This work extends previous work that investigated postulates of defeasible reasoning [10] and belief change [11] with human reasoners via surveys. Furthermore, this work is an extended abstracted of a paper that is currently under review for a special issue of the Journal of Applied Logic.

2. Background

2.1. Belief Revision

The first form of belief change we investigated was revision. It is an approach to reasoning with changing beliefs under the assumption that the world did not undergo a fundamental change. It is characterised by a belief set $\mathcal{K}$, a revision operation $*$ and reasoning rules referred to as postulates. A belief set is a set of propositional formulas closed under logical consequence. A revision operation allows a reasoner to add new information to his beliefs if the new information is consistent with his beliefs. A revision operation also allows a reasoner to add an exception to his beliefs to account for the situation where this exception or new information is inconsistent with his beliefs. Moreover, the result of a revision operation must always be that a reasoner's beliefs do not contradict

NMR'22: 20th International Workshop on Non-Monotonic Reasoning, August 07-09, 2022, Haifa, Israel

*Corresponding author.

bkrcla003@myuct.ac.za (C. K. Baker); tmeyer@cs.uct.ac.za (T. Meyer)

https://tinyurl.com/5n6vuyp7 (C. K. Baker); https://tinyurl.com/5ejbf2vr (T. Meyer)

0000-0002-3157-9989 (C. K. Baker); 0000-0003-2204-6969 (T. Meyer)

one another. There are eight postulates in the AGM [8] belief revision framework. (R1)–(R6) correspond to the core rationality postulates and the (R7)–(R8) correspond to supplementary postulates.

(R1) $\mathcal{K} * \mu$ implies μ
(R2) If $\mathcal{K} \wedge \mu$ is satisfiable, then $\mathcal{K} * \mu \equiv \mathcal{K} \wedge \mu$
(R3) If μ is satisfiable, then $\mathcal{K} * \mu$ is also satisfiable
(R4) If $\mathcal{K}_1 \equiv \mathcal{K}_2$ and $\mu_1 \equiv \mu_2$, then $\mathcal{K}_1 * \mu_1 \equiv \mathcal{K}_2 * \mu_2$
(R5) $(\mathcal{K} * \mu) \wedge \phi$ implies $\mathcal{K} * (\mu \wedge \phi)$
(R6) If $(\mathcal{K} * \mu) \wedge \phi$ is satisfiable, then $\mathcal{K} * (\mu \wedge \phi)$ implies $(\mathcal{K} * \mu) \wedge \phi$
(R7) If $\mathcal{K} * \mu_1$ implies μ_2 and $\mathcal{K} * \mu_2$ implies μ_1, then $\mathcal{K} * \mu_1$ is equivalent to $\mathcal{K} * \mu_2$
(R8) $(\mathcal{K} * \mu_1) \wedge (\mathcal{K} * \mu_2)$ implies $\mathcal{K} * (\mu_1 \vee \mu_2)$

2.2. Belief Update

The next form of belief change we investigated was update. It is an approach to reasoning with changing beliefs after some fundamental shift in the world occurred. It is characterised by a belief set $\mathcal{K}$, an update operation $\diamond$ and postulates for reasoning. As with revision, $\mathcal{K}$ refers to a logically closed set of propositional formulas. When we update $\mathcal{K}$ with new information μ, we are saying that we used to believe $\mathcal{K}$, we know now that μ holds, and we need to modify $\mathcal{K}$ by adding μ, acknowledging that we may have been wrong if μ contradicts $\mathcal{K}$. There are nine postulates in the KM [9] belief update framework.

(U1) $\mathcal{K} \diamond \mu$ implies μ
(U2) If $\mathcal{K}$ implies μ then $\mathcal{K} \diamond \mu$ is equivalent to $\mathcal{K}$
(U3) If both $\mathcal{K}$ and μ are satisfiable then $\mathcal{K} \diamond \mu$ is also satisfiable
(U4) If $\mathcal{K}_1 \leftrightarrow \mathcal{K}_2$ and $\mu_1 \leftrightarrow \mu_2$ then $\mathcal{K}_1 \diamond \mu_1 \leftrightarrow \mathcal{K}_2 \diamond \mu_2$
(U5) $(\mathcal{K} \diamond \mu) \wedge \phi$ implies $\mathcal{K} \diamond (\mu \wedge \phi)$
(U6) If $\mathcal{K} \diamond \mu_1$ implies μ_2 and $\mathcal{K} \diamond \mu_2$ implies μ_1 then $\mathcal{K} \diamond \mu_1 \leftrightarrow \mathcal{K} \diamond \mu_2$
(U7) If $\mathcal{K}$ is complete then $(\mathcal{K} \diamond \mu_1) \wedge (\mathcal{K} \diamond \mu_2)$ implies $\mathcal{K} \diamond (\mu_1 \vee \mu_2)$
(U8) $(\mathcal{K}_1 \vee \mathcal{K}_2) \diamond \mu \leftrightarrow (\mathcal{K}_1 \diamond \mu) \vee (\mathcal{K}_2 \diamond \mu)$
(U9) If $\mathcal{K}$ is complete and $(\mathcal{K} \diamond \mu) \wedge \phi$ is satisfiable then $\mathcal{K} \diamond (\mu \wedge \phi)$ implies $(\mathcal{K} \diamond \mu) \wedge \phi$

Revision and update differ from non-monotonic logic using the concept of orders on interpretations. A homogeneous relation $\leq$ on some given set P, so that by definition $\leq$ is some subset of $P \times P$ and the notation $a \leq b$ is used in place of $(a, b) \in P$, is called a preorder if the relation is also transitive and reflexive. A reflexive relation has the property that $a \leq a$ for all $a \in P$. A transitive relation has the property that if $a \leq b$ and $b \leq c$ then $a \leq c$ for all $a, b, c \in P$. If a preorder is also anti-symmetric, that is, $a \leq b$ and $b \leq a$ implies $a = b$, then it is a partial preorder. A preorder is total if $a \leq b$ or $b \leq a$ for all $a, b \in P$. A revision operator satisfies postulates (R1)–(R6) using the notion of a total preorder on interpretations while an update operator satisfies postulates (U1)–(U6) using the notion of a partial preorder on interpretations. By replacing postulates (U6) and (U7) with a new postulate (U9), the class of update operators can be designed using total preorders. The second and more important difference between revision and update is that, in the case of update, a different ordering is induced by each model of $\mathcal{K}$, while for revision, only one ordering is induced by the whole of $\mathcal{K}$.

3. Methodology

Our empirical investigation took place through four experiments. In the first experiment, we prepared a survey of 30 general statements about the world for participants to evaluate for clarity and bias. 7 participants had to complete a table in which they identified statements with ambiguous language and biased examples. In the second experiment, we prepared a survey of 30 general statements about the world taken from refining the material in the first experiment. 30 participants evaluated the degree to which they believed each of the statements in the survey and explained their answers. In the third experiment, we prepared a survey of English statements corresponding to translations of the AGM postulates for belief revision. 35 participants on Mechanical Turk (MTurk) evaluated the degree to which they believed each statement in the survey. We tested our hypothesis statistically and determined whether the association between the premises and the conclusion for each postulate holds for the general English-speaking reasoner. In the last experiment, we used the same material from the belief revision experiment to instantiate the KM belief update postulates. The experimental setup followed a similar approach to the belief revision experiment. We obtained ethical clearance from the Faculty of Science Ethics Research Committee at the University of Cape Town. We include the consent forms and a link to our data management plan in our Github project repository, linked in Appendix A. For the bulk of our reasoning experiments, we used Google Forms to design our surveys, and we used Mechanical Turk to crowdsource our data collection.

4. Results

In this work, we investigated the endorsements of each component of the AGM postulates when formulated as material implication statements. We found evidence for whether or not the participants found our concrete instantiations of the AGM postulates plausible. We determined whether the postulates hold in general. The results show that the participants' reasoning tends to be consistent with the 9 AGM postulates, with the significance of the association between the endorsement of the premises and the endorsement of the conclusion ranging from non-significant to highly significant. The number of logical violations per postulate is generally low with a range of 0 to 4 violations ($< 12\%$ of participants). The exception is postulate (R5) with 54,29% (19 participants) endorsing the premises, but not the conclusion.

We also investigated the endorsements of each component of the KM postulates when formulated as material implication statements. We found evidence for whether or not the participants found our concrete instantiations of the KM postulates plausible. We determined whether the postulates hold in general. The results show that the participants' reasoning tends to be consistent with the 9 postulates of KM belief update, with the significance of the association between the endorsement of the premises and the endorsement of the conclusion ranging from non-significant to highly significant. The number of logical violations per postulate is generally low as well with a range of 0 to 8 violations ($< 23\%$ of participants). The exception is postulate (U8) with 48,57% (17 participants) endorsing the premises, but not the conclusion.

5. Conclusions and Future Work

Our work builds on previous empirical studies involving human subjects who are tasked with reasoning nonmonotonically. We created a reproducible approach for empirically investigating the plausibility of postulates of belief change. This approach accounts for the effect of the premises and conclusion of each postulate and determines whether the overall postulate is found plausible with statistically significant evidence. We applied this approach to the formal theory of belief revision and update. We hypothesised that human belief change is consistent with the AGM postulates of belief revision and the KM postulates of belief update. The results show that the participants' reasoning tends to be consistent with the 8 AGM postulates, (R1)–(R8), with the significance of the association between the endorsement of the premises and the endorsement of the conclusion ranging from non-significant to highly significant. The results also show that the participants' reasoning tends to be consistent with the 9 KM postulates, (U1)–(U9), with the significance of the association between the endorsement of the premises and the endorsement of the conclusion ranging from non-significant to highly significant.

In future work, we will refine our approach in the following way. We will conduct a theoretical and empirical investigation of the postulates of belief revision and update. The theoretical part will build on this work by exploring inter-postulate and inter-framework relationships. The empirical part will focus on identifying representations of the postulates that support both the theory and the beliefs of human reasoners. It will involve the development of an online reasoning tool that automates the production and presentation of structured reasoning examples in a survey setting. The structured reasoning examples will depict the static and dynamic nature of changing beliefs in terms of a revision and an update, respectively. In turn, the tailored surveys created using the reasoning tool can be used to elicit responses from human reasoners, the design of which improves upon the limitations of question types from conventional survey platforms like Google Forms and Microsoft Forms. Furthermore, non-parameterised statistical methods, that is, methods that do not assume how the sample data is distributed, e.g. the Wilcoxon signed-rank test [12, 13], will be used to interpret the significance of the postulates of belief change, as found by human reasoners.

Acknowledgments

We wish to express our sincere gratitude and appreciation to the DSI – CSIR Interbursary Support (IBS) Programme and the Centre for Artificial Intelligence Research (CAIR) for financial support.

References

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

[2] M. Ragni, C. Eichhorn, T. Bock, G. Kern-Isberner, A. Tse, Formal nonmonotonic theories and properties of human defeasible reasoning, Minds and Machines 27 (2017) 79–117. doi:10.1007/s11023-016-9414-1.

[3] R. M. Byrne, Suppressing valid inferences with conditionals, Cognition 31 (1989) 61–83.

[4] F. Lévy, A survey of belief revision and updating in classical logic, International Journal of Intelligent Systems 9 (1994) 29–59.

[5] D. Dubois, H. Prade, A survey of belief revision and updating rules in various uncertainty models, International Journal of Intelligent Systems 9 (1994) 61–100.

[6] C. Boutilier, N. Friedman, J. Y. Halpern, Belief revision with unreliable observations, in: AAAI/IAAI, 1998, pp. 127–134.

[7] S. Doutre, J.-G. Mailly, Constraints and changes: A survey of abstract argumentation dynamics, Argument & Computation 9 (2018) 223–248.

[8] C. E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change: Partial meet contraction

and revision functions, Journal of Symbolic Logic 50 (1985) 510–530. doi:10.2307/2274239.
[9] H. Katsuno, A. O. Mendelzon, On the difference between updating a knowledge base and revising it, Belief revision (1991) 183.
[10] C. Baker, C. Denny, P. Freund, T. Meyer, Cognitive defeasible reasoning: the extent to which forms of defeasible reasoning correspond with human reasoning, in: Proceedings of the First Southern African Conference for Artificial Intelligence Research (SACAIR 2020), CCIS, Springer, 2020, pp. 119–219.
[11] C. Baker, T. Meyer, Belief change in human reasoning: An empirical investigation on mturk, in: Proceedings of the Second Southern African Conference for Artificial Intelligence Research (SACAIR 2021), 2021, pp. 520–536.
[12] X. Li, Y. Wu, M. Wei, Y. Guo, Z. Yu, H. Wang, Z. Li, H. Fan, A novel index of functional connectivity: phase lag based on wilcoxon signed rank test, Cognitive Neurodynamics 15 (2021) 621–636.
[13] L. Zhou, Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods, Knowledge-Based Systems 41 (2013) 16–25.

A. Additional resources

The Github repository for this work, containing supplementary material and code scripts, can be accessed via this URL, https://tinyurl.com/2p98m76n.