Supplementary material: Asking human reasoners to judge postulates of belief change for plausibility*

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. 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. 50 participants were recruited on Mechanical Turk (MTurk) to evaluate 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.

1.1. Our representation of the postulates

We represented the AGM and KM postulates as material implication rules, e.g. premises $\rightarrow$ conclusion. This allowed us to analyse the influence of each component, the premises and the conclusion, on the plausibility of the postulate. This also allowed us to determine whether participants committed logical violations by comparing the endorsements of each rule to its corresponding interpretation in propositional logic. (R9) was omitted from our investigation, since it is a pointwise revision operator which did not pertain to our study. We give the component rules for the AGM and KM postulates, and describe our methods of analysis, in our project repository linked in Appendix A.

1.2. Possibility theory as a method of analysis

We present possibility theory, introduced by Zadeh [1] and refined by Dubois and Prade [2], as an approach for validating the responses in our experiments surveying the belief revision and update postulates. In particular, we conducted experiments (refer to Experiment 3 and 4) in which these postulates were ranked by participants in terms of how plausible it was to them. The postulates were represented in a manner that enabled comparison with the interpretation of default rules in possibility theory. For context, we recall that Benferhat et al. [3] and Dubois and Prade [4] applied possibility theory to study postulates of non-monotonic reasoning. Benferhat et al. [5] showed that the possibilistic approach contributed a faithful representation of the KLM [6] postulates. Thus, a default rule "if a then b, generally", denoted $a \mid\sim b$, is represented by the constraint,

$$\Pi(\alpha \wedge \beta) > \Pi(\alpha \wedge \neg\beta)$$

on a possibility measure Π describing the semantics of the available knowledge. The constraint says that in the con-

NMR'22: 20th International Workshop on Nonmonotonic Reasoning Workshop, 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)

text where α is true, there exists situations where having β true is strictly more plausible than any situations where β is false in the same context. Participants in our experiments judged each component of the postulate separately. To enable comparison with the interpretation of defaults in possibility theory, we represented each material implication statement as a default rule, "generally from the premises, deduce the conclusion". In our computation of Π for each default rule, we used frequency counts obtained in the data analysis step of our experiments. The counts were tabulated in a 2×2 contingency table format showing combinations of positive and negative participant endorsements of the premises and the conclusion. For example, suppose the premises of AGM postulate R1 is the formula α and its conclusion is β. Then, the corresponding contingency table is indexed to find the combinations of α and β needed for Π. This means $\alpha \wedge \beta$ is the frequency of the endorsement of both the premises and the conclusion, while $\alpha \wedge \neg\beta$ is the frequency of the endorsement of the premises and the frequency of the non-endorsement of the conclusion. A default rule was found plausible by English-speaking human reasoners in general if the condition $\Pi(\alpha \wedge \beta) > \Pi(\alpha \wedge \neg\beta)$ was satisfied.

1.3. Hypotheses testing

We followed a quantitative approach to test our hypothesis that human reasoning is consistent with the AGM postulates of belief revision. The task in experiment 3 was for participants to rank concrete natural language rules corresponding to the AGM postulates of belief revision. Each postulate was first decomposed into material implication rules to obtain a premise and conclusion. Participants had to rank the premises separately from the conclusion. No explanation was required. The metric used was plausibility, a subjective measure, determined by participants, of whether the postulate premises and conclusion hold in everyday reasoning. The responses were indicated on a linear scale from 1 (implausible) to 10 (extremely plausible). Participants were recruited anonymously via MTurk. Participants who consented to this experiment were asked to complete a survey on Google Forms. The average rank for each rule corresponding to the premises and the conclusion was computed. An average rank of 6 or greater indicated agreement with the rule. An average rank of 5 or less indicated disagreement with the rule. We computed a contingency table containing the frequencies of endorsements of the premises and the conclusion of each postulate. To test whether there is a relationship between the endorsement of the premises and the endorsement of the conclusion, we conducted a statistical hypothesis test. We measured the linear association between the premises and conclusion using the phi-coefficient. The phi-coefficient formula for the 2×2 contingency table is:

$$\phi = \frac{AD - BC}{\sqrt{((A+B)(C+D)(A+C)(B+D))}} \quad (1)$$

The phi-coefficient produces a value between 0 and 1, extremes which denote no association and perfect association respectively. The closer the score to 0, the weaker the association between the premises and conclusion. The closer the score to 1, the stronger the association between the premises and conclusion. An error occurs when the denominator of the phi-coefficient evaluates to 0, because of any of the four sums in the denominator evaluating to 0. This is treated by setting the denominator to 1, an arbitrary value [7]. The result is a phi-coefficient of 0 which is means that no association could be determined. Once computed, the phi-coefficient was tested for statistical significance using three tests: chi-squared, chi-squared with Yates' correction and Fisher's exact test. We set the significance level to 5% for each test. A comparison of all three tests together provides us with a stronger position to assert the significance of the phi-coefficient than any of the tests alone. As the postulates for belief revision intend to capture belief change in human reasoning, we assume that participants' reasoning will be consistent with it. However, we expected to find diverse responses and discussed trends and inconsistencies in our sample. We generalised whether the postulates for belief revision are consistent with English-speaking human reasoners. We include a comparison of each participant's response to the average response for the rule, and discussion of the influence of age and gender on the rank of each rule in our project repository. Further, we used the framework of possibility theory [2] by interpreting the material implication statements as default rules, to evaluate whether the association for each postulate is significant in general, or whether the significance was specific to the concrete instances. Experiment 4 tested translations of the KM postulates of belief update for plausibility in an analogous manner to experiment 3. We include a comparison of each participant's response to the average response for the rule, and discussion of the influence of age and gender on the rank of each rule in our project repository.

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] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy sets and systems 1 (1978) 3–28.

[2] D. Dubois, H. Prade, Possibilistic logic - an overview, in: Handbook of the History or Logic, volume 9, 2014, pp. 197–255.
[3] S. Benferhat, D. Dubois, H. Prade, Practical handling of exception-tainted rules and independence information in possibilistic logic, Applied intelligence 9 (1998) 101–127.
[4] D. Dubois, H. Prade, Non-monotonic reasoning and uncertainty theories, Nonmonotonic Reasoning. Essays Celebrating its 30th Anniversary.. Gerhard Brewka, Victor Marek, Miroslaw Truszczynski (Eds.), College Publications (2011) 141–176.
[5] S. Benferhat, D. Dubois, H. Prade, Nonmonotonic reasoning, conditional objects and possibility theory, Artificial Intelligence 92 (1997) 259–276.
[6] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1990) 167–207.
[7] E. Souza, T. Negri, First prospects in a new approach for structure monitoring from gps multipath effect and wavelet spectrum, Advances in Space Research 59 (2017) 2536–2547.

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.