Proposal: An empirical investigation of the extent to which human belief change is consistent with AGM belief revision and KM belief update

Clayton Baker
University of Cape Town and CAIR
bkrcla003@myuct.ac.za

CCS CONCEPTS

• **Computing methodologies → Nonmonotonic, default reasoning and belief revision**; *Reasoning about belief and knowledge*; • **Theory of computation** → Logic.

KEYWORDS

non-monotonic reasoning, belief revision, belief update, belief change, propositional logic, survey, Mechanical Turk

1 SUPPLEMENTARY INFORMATION

1.1 Test for consistency

We outline the test of consistence of the spreading activation process with AGM belief revision in the following:

(1) For each postulate of AGM belief revision, we first decompose them into its LP and RP components.

(2) We will choose a set of concrete rules that match the LP and RP components involved in the postulate. Each rule must be such that some people endorse it as plausible and some other people do not. By assuming that these postulates are not target properties of belief revision, but rather that they are emergent ones, we do not include the symbolic representations and use concrete material to implant the information into memory. In addition, we assume that a single concrete instance should suffice in order to test a given postulate if "the concrete material is such that no implicit knowledge is available or if the instances of the rule refer explicitly to the most typical one". It is important the the influence of material on endorsement of the postulates be tested.

(3) Participants then independently rate the degrees of possibility for $LP \wedge RP$ and $LP \wedge \neg RP$. It allows to evaluate the plausibility of every rule involved in the postulates of AGM belief revision, for each participant, according to the semantics of possibilistic inference. A simple semantics for plausible rules has been proposed by Benferhat, Dubois and Prade [1] and Dubois and Prade [4] within the framework of Possibility theory. According to this semantics, a conditional assertion $\alpha \mid\sim \beta$ can be viewed as a constraint expressing that a situation where α and β are both true has a greater possibility (Π) than a situation where α is true and β is false. Formally, $\alpha \mid\sim \beta$ is a plausible rule if $\Pi(\alpha \wedge \beta) \geq \Pi(\alpha \wedge \neg\beta)$.

(4) For each postulate, (i) a contingency table is constructed and (ii) the coefficient of association between the variables "endorsement of LP"and "endorsement of RP", and the differences between cells A and B, cells A and C, and cells B and C are computed. The contingency table for step (i) is partitioned into four cells A, B, C and D. Cell A refers to those who endorse LP and RP, Cell B refers to those who endorse LP but not RP, Cell C refers to those who do not endorse LP but endorse RP and Cell D refers to those who endorse neither LP nor RP.

(5) The consistency of human inference with an AGM postulate is checked according to a set of decision rules (refer to Table 1). In the case of decision rule 3*, the inference "RP $\longrightarrow$ LP" contributes also to the association degree but less than the "LP $\longrightarrow$ RP" inference. We apply this same approach, steps (1)-(5), for testing KM belief update.

Table 1: Decision rules for testing consistency

Rule no.	Rule description
1.	If (LP and RP are not significantly associated) or (A is not significantly greater than B, with B close to 0) then (the property under consideration is not corroborated)
2.	If (LP and RP are significantly positively associated) and (A is significantly greater than B, with B close to 0) and (A is not significantly different from C) then (the property under consideration is corroborated)
3*.	If (LP and RP are significantly positively associated) and (A is significantly greater than B, with B close to 0) and (A is significantly different from C) (C is significantly different from B) then (the property under consideration is corroborated)

1.2 Experimental testing of KM belief update

Using a similar methodology to the experiment testing AGM belief revision, we will conduct an experiment testing KM belief update. We mention some important details below, but refer the reader to the proposal document for the complete procedure.

Participants. We will recruit 50 native English speakers as participants through Mechanical Turk. No participant should have formal training in logic. Participants in this experiment should not have taken part in the pre-experiment nor the experiment testing AGM belief revision.

Material. KM belief update includes nine postulates: *U1*, *U2*, *U3*, *U4*, *U5*, *U6*, *U7*, *U8* and *U9*. Three of the postulates (*U1*, *U5* and *U8*) have two rules: one rule in LP and one rule in RP. Two of the postulates (*U3* and *U4*) have three rules: two rules in LP and one rule in RP. Two of the postulates (*U2* and *U7*) have four rules: two rules in LP and two rules in RP. One of the postulates (*6*) have six rules: three rules in LP and three rules in RP. One of the postulates (*U9*) has five rules: three rules in RP and two rules in LP. In total there are thirty-one rules, consisting of seventeen in LP and fourteen in RP. The seventeen LP rules will be derived from the experiment. The fourteen RP rules will be derived according to target postulates. The status these thirty-one rules have for each participant will be checked by means of two questions.

Design and Procedure. The sixty-two questions will be presented in random order in the style of a survey, as before.

Predictions. The encoding of answers will be the same as in the pre-experiment. We follow a similar approach to AGM belief revision in terms of the hypothesis predictions and testing.

Results. We will compute ELP and ERP rates for each postulate of KM belief update and produce contingency tables for each postulate, in the style of the pre-experiment.

1.3 Participant remuneration

We will use an exchange rate of R17 = $1. The minimum wage in the USA is $7,25/hour (R123,25/hour). The South African minimum wage is R15,57/hour. We use the average minimum wage per hour calculated as follows:

$$\frac{\text{R}123,25 + \text{R}15,57}{2} = \text{R}69,40 \tag{1}$$

We will use the amount in Equation (1) as the reward paid to our participants because we expect that our survey will be completed within an hour. To avoid paying further costs, we will conduct the surveys on Mechanical Turk in batches of no more than 9 participants. At 20% of R69,40, the fee owed to Mechanical Turk for an individual task is R13,88. The cost of an individual survey on Mechanical Turk thus comprises the sum of the reward paid to Workers and the fee owed to Mechanical Turk:

$$\text{R}69,40 + \text{R}13,88 = \text{R}83,23 \tag{2}$$

The pre-experiment requires 30 participants and both experiments 1 and 2 require 50 participants. The total cost for the remuneration of participants in our experiments is calculated as follows:

$$(30 \text{ x R}83,28) + (50 \text{ x R}83,28) + (50 \text{ x R}83,28) = \text{R}10826,40 \tag{3}$$

1.4 Defeasible reasoning

In the seminal work by Kraus, Lehmann and Magidor (KLM) [9], they studied general patterns of non-monotonic reasoning and attempted to isolate properties which map non-monotonic reasoning to positive properties, rather than negatively describing non-monotonic reasoning by the property it does not enjoy, monotonicity. They define five families of non-monotonic consequence relations (Systems C, CL, P, CM and M) and provide representation theorems for each. An outcome of this paper is the defacto standard in non-monotonic reasoning, System P, which computes a preferential consequence relation based on the KLM properties Reflexivity, Left Logical Equivalence, Right Weakening, Cautious Monotonicity, CUT and Or. Gärdenfors and Makinson [7] give an elegant interpretation of non-monotonic inferences in terms of underlying *expectations*. They provide two ways of modelling expectations: by expectation sets equipped with selection functions, and by expectation relations. They provide representation theorems and establish relations with preferential models. Using the notion of an *expectation*, Gärdenfors and Makinson argue that belief revision and non-monotonic inference relations can be treated uniformly by viewing the relation of epistemic entrenchment used in Gärdenfors [5] and Gärdenfors and Makinson [6] as a kind of expectation ordering. Lehmann [10] provides a logic of normal defaults, different to the one proposed by Reiter [13], which is rich enough not to require the consideration of non-normal defaults. In his work, Lehmann defines the lexographic closure of any given finite set $\mathcal{D}$ of defaults, and provides examples to show that the lexicographic closure defined corresponds to the basic intuition behind Reiter's defaults. In a subsequent work, Lehmann [11] studies stereotypical reasoning as a basic form of non-monotonic reasoning. Lehmann proposes a formal model for stereotypical reasoning and examines its logical properties. Under weak assumptions, Lehmann finds that stereotypical reasoning is cumulative. Lin and Chen [12] proposed a restricted semantics of four-valued logic for default reasoning to resolve the problems of inconsistency and incoherence whilst retaining the classical extensions in the presence of consistency and coherency. Their results prove that the restricted semantics can maintain both the expressiveness and reasoning ability of default logic and an approach is provided to compute the restricted extensions by reducing them to classical ones. Britz and Varzinczak [3] investigated defeasible modes of reasoning and introduce new modal operators with which to formalise the notion of defeasible necessity and distinct possibility. In their work, Britz and Varzinczak [3] demonstrate how KLM-style conditionals can smoothly be integrated with their richer language. Britz *et al.* [2] present an approach to defeasible reasoning for the description logic $\mathcal{ALC}$. Their results are based on the work done by Kraus, Lehmann and Magidor (KLM) on defeasible conditionals in the propositional case. Britz *et al.* [2] study versions of a preferential semantics of two forms of defeasible subsumption, and link these constructions formally to KLM-style syntactic properties via representation results. Additionally, Britz *et al.* [2] investigated an appropriate form of *defeasible entailment* for the enriched version of $\mathcal{ALC}$ and their results include an algorithm for the computation of a form of defeasible entailment known as *rational closure* in the propositional case. Governatori *et al.* [8] investigated how to change a preference relation in a non-monotonic logic in order to change the conclusions of the theory itself. The main result of their work is the proof that the problem of revising a non-monotonic theory by changing only the superiority order between conflicting rules is generally computationally hard.

REFERENCES

[1] S. Benferhat, D. Dubois, and H. Prade. 1992. Representing default rules in possibilistic logic.. In *Proceedings of 3rd International conference on principles of knowledge representation and reasoning (KR'92)*. 673–684.

[2] K. Britz, G. Casini, T. Meyer, and I. Varzinczak. 2019. A KLM perspective on defeasible reasoning for description logics. *Description Logic, Theory Combination, and All That* (2019), 143–173.

[3] K. Britz and I. Varzinzak. 2018. From KLM-style conditionals to defeasible modalities, and back. *Journal of Applied Non-Classical Logics* 28, 1 (2018), 92–121.

[4] D. Dubois and H. Prade. 1996. *Conditional objects, possibility theory and default rules*. 301–336.

[5] P. Gärdenfors. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, Massachusetts.

[6] P. Gärdenfors and D. Makinson. 1988. Revisions of Knowledge Systems Using Epistemic Entrenchment. In *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning About Knowledge (TARK '88)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 83–95. http://dl.acm.org/citation.cfm?id=1029718.1029726

[7] P. Gärdenfors and D. Makinson. 1994. Nonmonotonic inference based on expectations. *Artificial Intelligence* 65, 2 (1994), 197–245.

[8] G. Governatori, F. Olivieri, M. Cristani, and S. Scannapieco. 2019. Revision of defeasible preferences. *International Journal of Approximate Reasoning* 104 (2019), 205–230.

[9] S. Kraus, D. Lehmann, and M. Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.* 44 (1990), 167–207.

[10] D. Lehmann. 1995. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence* 15, 1 (1995), 61–82.

[11] D. Lehmann. 1998. Stereotypical reasoning: logical properties. *Logic Journal of IGPL* 6, 1 (1998), 49–58.

[12] Z. Lin and C. Chen. 2018. Restricted semantics for default reasoning. *Annals of Mathematics and Artificial Intelligence* 84, 3-4 (2018), 185–200.

[13] R. Reiter. 1980. A Logic for Default Reasoning. *Artificial Intelligence* 13 (04 1980), 81–132. https://doi.org/10.1016/0004-3702(80)90014-4