Supplementary Material for "Belief Change in Human Reasoning: An Empirical Investigation on MTurk" ⋆

Clayton Kevin Baker[1] [0000−0002−3157−9989] and Thomas Meyer[1] [0000−0003−2204−6969]

[1]University of Cape Town and Centre for Artificial Intelligence Research (CAIR), Cape Town, South Africa
clayton.baker@uct.ac.za
tmeyer@cair.org.za

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic **50**, 510–530 (1985). https://doi.org/10.2307/2274239 2, 6
2. Andreotta, M., Nugroho, R., Hurlstone, M., Boschetti, F., Farrel, S., Walker, I., Paris, C.: Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis. Behavior research methods **51**(4), 1766–1781 (2019) 7
3. Bentley, F., Daskalova, N., White, B.: Comparing the reliability of amazon mechanical turk and survey monkey to traditional market research surveys. In: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1092–1099 (2017) 7
4. Creswell, J.W.: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE Publications, Inc., 4th edn. (January 2014) 7
5. Dubois, D., Prade, H.: Fuzzy sets and systems: theory and applications, vol. 144. Academic press (1980) 3, 4
6. Dubois, D., Prade, H.: Possibilistic logic - an overview. In: Handbook of the History or Logic. vol. 9, pp. 197–255 (2014) 3
7. Fermé, E., Hansson, S.O.: Belief Change: Introduction and Overview. SpringerBriefs in Intelligent Systems, Springer International Publishing AG, Cham (2018) 5
8. Grootswagers, T.: A primer on running human behavioural experiments online. Behavior research methods pp. 1–4 (2020) 7
9. Katsuno, H., Mendelzon, A.O.: On the difference between updating a knowledge base and revising it. Belief revision p. 183 (1991) 2, 6
10. Machleid, F., Kaczmarczyk, R., Johann, D., Balčiūnas, J., Atienza-Carbonell, B., von Maltzahn, F., Mosch, L.: Perceptions of digital health education among european medical students: Mixed methods survey. Journal of medical Internet research. **22**(8) (2020-8-14) 7

⋆ 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.

11. Makinson, D.: Bridges from classical to nonmonotonic logic. Texts in computing ; v. 5, King's College, London (2005) 4, 5
12. Nadarzynski, T., Miles, O., Cowie, A., Ridge, D.: Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study. Digital health **5**, 2055207619871808 (2019) 7
13. O'Halloran, K., Tan, S., Pham, D., Bateman, J., Vande Moere, A.: A digital mixed methods research design: Integrating multimodal analysis with data mining and information visualization for big data analytics. Journal of Mixed Methods Research **12**(1), 11–30 (2018) 7
14. Paolacci, G., Chandler, J.: Inside the turk: Understanding mechanical turk as a participant pool. Current Directions in Psychological Science **25**, 184–188 (2014) 7
15. Pollock, J.: Defeasible reasoning. Cognitive Science **11**(4), 481–518 (1987). https://doi.org/https://doi.org/10.1016/S0364-0213(87)80017-4, `http://www.sciencedirect.com/science/article/pii/S0364021387800174` 5
16. Springer, V., Martini, P., Lindsey, S., Vezich, I.: Practice-based considerations for using multi-stage survey design to reach special populations on amazon's mechanical turk. Survey Practice **9**(5), 1–8 (2016) 7
17. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and systems **1**(1), 3–28 (1978) 3, 4

A Background

In the following, we present background on the topics of propositional logic, nonmonotonic reasoning and belief change before concluding with a description of our chosen belief change frameworks, namely AGM [1] belief revision and KM [9] belief update.

A.1 Propositional Logic

Propositional logic deals with statements called propositions. Examples of propositions are "Kyle drives a blue car", "It is sunny", and "The ice-cream shop is closed". These propositions are called atoms because they cannot be decomposed into smaller propositions without giving up the original meaning. A set of atoms, $\mathcal{P} = \{\mathbf{p}, \mathbf{q}, ..\}$, abbreviates a set of facts about a system. We can build more complex formulas, called sentences, from the propositional atoms, using unary (negation) and binary (conjunction, disjunction, material conditional and material biconditional) operators. If α is a propositional formula, then α can be defined as follows:

$$\alpha ::= \mathbf{p} \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \tag{1}$$

The precedence associated with these operators in descending order is: $\neg$, $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$. The propositional language $\mathcal{L}$ generated by $\mathcal{P}$ is the set of formulas defined recursively as follows:

- If $\alpha \in \mathcal{P}$ then α is a formula of $\mathcal{L}$;

– If α and β are formulas of $\mathcal{L}$, then so are $(\neg\alpha)$, $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, $(\alpha \rightarrow \beta)$ and $(\alpha \leftrightarrow \beta)$.

In the following, we introduce notions of a propositional interpretation and the satisfiability of propositional formulas.

Definition 1. *(interpretation) An interpretation I is a total function that assigns one of the truth-values T or F to every atom in $\mathcal{P}$. Formally, I is given by $I : \mathcal{P} \longrightarrow \{T, F\}$. A set of interpretations, also referred to as states or possible worlds, is denoted by W.*

We make use of interpretations in our analysis of the experiment material.

A.2 Possibility Theory

Possibility theory, or possibilistic logic, is a weighted logic that qualitatively handles uncertainty by associating certainty, or priority levels, to classical logic formulas [6]. The basic building blocks of possibility theory originate from the work of Zadeh [17] and have been studied in detail by Dubois and Prade [5, 6]. Zadeh starts from the idea of a possibility distribution, to which he associates a possibility measure. In the definition that follows, we let U be a set of states of affairs (or descriptions thereof), or states for short. This set can be the domain of an attribute (numerical or categorical), the Cartesian product of attribute domains, the set of interpretations of a propositional language, etc.

Definition 2. *(Possibility distributions) A possibility distribution is a mapping π from U to a totally ordered scale $\mathcal{S}$, with top denoted by 1 and bottom by 0. In the finite case, $\mathcal{S} = \{1 = \lambda_1 > ... > \lambda_n > \lambda_{n+1} = 0\}$.*

The possibility scale can be the unit interval as suggested by Zadeh, or generally any finite chain, or even the set of non-negative numbers, in which case the conventions are opposite: 0 means possible and ∞ means impossible. The function π represents the state of the knowledge of an agent (about the actual states of affairs), also called an epistemic state, distinguishing what is plausible from what is less plausible, what is the normal course of things and what is not, what is surprising from what is expected. It represents a flexible restriction on what is the actual state of facts with the following conventions:

– $\pi(u) = 0$ means that possibility state u is rejected as impossible;
– $\pi(u) = 1$ means that possibility state u is totally possible (= plausible).

The larger $\pi(u)$, the more possible, i.e., plausible the state u is. Formally, the mapping π is the membership function of a fuzzy set [17] where membership grades are interpreted as plausibility. If the universe U is exhaustive, at least one of the elements in $\mathcal{S}$ should be the actual world, so that $\exists u, \pi(u) = 1$ (normalisation). This condition expresses the consistency of the epistemic state described by π. In the {0,1}-valued case, π is just the characteristic function of a subset $E \subseteq U$ of mutually exclusive states, ruling out all those states outside of E considered as impossible. Our experiment material can be viewed on the spectrum of a possibility distribution from 0 to 1. We use this scale as a means to judge the plausibility of the reasoning material in our experiments. Given a simple query of the form "does event A occur?" or "is the corresponding proposition a true?", where A is the subset of states, the set of models of a, the response to the query can be obtained by computing the degrees of possibility of A [17] and its complement A^C:

$$\Pi(A) = \sup_{u \in A} \pi(u); \Pi(A^C) = \sup_{s \notin A} \pi(u); \tag{2}$$

$\Pi(A)$ evaluates to what extent A is consistent with π, while $\Pi(A^C)$ can be easily related to the idea of certainty of A. Indeed, the less $\Pi(A^C)$ the more A^C is impossible and the more certain A is. If the possibility scale $\mathcal{S}$ is equipped with an order-reversing map denoted by $\lambda \in \mathcal{S} \mapsto v(\lambda)$, it enables a degree of necessity (certainty) of A to be defined in the form $N(A) = v(\Pi(A^C))$, which expresses the well-known duality between possibility theory and necessity. $N(A)$ evaluates to what extent A is certainly implied by π. If $\mathcal{S}$ is the unit interval, then it is usual to choose $v(\lambda) = 1 - \lambda$, so that $N(A) = 1 - \Pi(A^C)$ [5]. Generally, $\Pi(U) = N(U) = 1$ and $\Pi(\emptyset) = N(\emptyset) = 0$ (since π is normalised to 1). In the {0,1}-valued case, the possibility distribution comes down to the disjunctive (epistemic) set $E \subseteq U$, and possibility and necessity are then defined as follows:

- $\Pi(A) = 1$ if $A \cap E \neq \emptyset$, and 0 otherwise: function Π checks whether A is logically consistent with the available information or not.
- $N(A) = 1$ if $E \subseteq A$, and 0 otherwise: function N checks whether A is logically entailed by the available information or not.

Possibility measures satisfy the characteristic "maxitivity" property:

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \tag{3}$$

Necessity measures satisfy an axiom dual to that of possibility measures:

$$N(A \cap B) = \min(N(A), N(B)) \tag{4}$$

A.3 Non-Monotonic Reasoning

In the context of Sherlock Holmes, a fictional English detective, drawing conclusions from available evidence, Makinson (2005) [11] described reasoning as an appeal "made not only to the observations explicitly mentioned, but also, implicitly, to a reservoir of background knowledge, a supply of rules of thumb, a range of heuristic guides". Most important, conclusions may be withdrawn as more information comes to hand, and new ones may be advanced in their place [11]. This does not necessarily mean that there was an error in the earlier reasoning [11]. The previous inferences may still be recognized as the most reasonable ones to have been made with the information available [11]. In formal terms, we are said to be reasoning non-monotonically when we allow that a conclusion that is well-drawn from given information may need to be withdrawn when we come into possession of further information, even when none old premises are abandoned [11]. Pollock [15] said that when one judges the colour of something based on how it looks to him, he is not reasoning deductively. Such reasoning is defeasible, in the sense that the premises taken by themselves may justify us in accepting the conclusion, but when additional information is added, that conclusion may no longer be justified [15]. In the non-monotonic and defeasible case, it is clear that the conclusion of an argument based on some premises may change, should the premises or external factors change. We discuss this change in terms of beliefs, next.

A.4 Belief Change

Fermé and Hansson [7] give an example of reasoning with belief change: they consider a set of sentences in natural language: "Juan was born in Puerto Carreño (α)", "José was born in Puerto Ayacucho (β)" and "Two people are compatriots if they were born in the same country (γ)". This set represents all the currently available information about Juan and José. Suppose that we receive the following piece of new information: "Juan and José are compatriots (δ)". If we add the new information to our corpus of beliefs, then we obtain a new set of beliefs that contains the sentences α, β, γ and δ [7]. We can define an operation of addition as one that takes a sentence and a set of previous beliefs and returns the minimal set that includes both the previous beliefs and the new sentence. This operation exemplifies the simplest way of changing a set of sentences [7]. In contrast, consider that upon consulting an atlas we discover to our surprise that Puerto Carreño is in Colombia (ϵ) and Puerto Ayacucho is in Venezuela (ϕ) [7]. If we add ϵ and ϕ to the set $\{\alpha, \beta, \gamma, \delta\}$, the result will be a set with contradictory information: Juan and José are compatriots but Puerto Carreño and Puerto Ayacucho do not belong to the same country. The addition does not necessarily reflect the notion of a consistent revision [7]. If we wish to retain consistency, then some subset of the original set must be discarded or perhaps a part of the new information has to be rejected [7]. In our example, there are several possible alternatives. The information about Juan's or José's birthplace could be wrong, and so could the atlas. Finally, the claim that Juan and José are compatriots could be wrong. Any of these three options, either individually or combined, will allow us to solve the problem of the incompatibility between the original and the new information or beliefs. Consequently, we can specify an operation that takes a set and a sentence and returns a new consistent set. The new set includes parts (or all) of the beliefs in the original set and it also includes the new sentence (if we are willing to accept it). The outcome of a revision can be expressed as a consistent subset of the outcome of the addition.

AGM Belief Revision Belief revision 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 K, a revision operation $*$ and reasoning rules referred to as postulates. A belief set is a set of propositional formulas closed under classical logical consequence (C_n). 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 one another. There are eight postulates in the AGM belief revision framework, described in [1] and shown in Table 1.

Table 1. AGM Postulates

R1. $K = C_n(K)$ and $K * p = C_n(K * p)$
R2. If $K * p \models \alpha$ then $K + p \models \alpha$
R3. If $K \not\models \neg p$ then (if $K + p \models \alpha$ then $K * p \models \alpha$)
R4. $p \in K * p$
R5. If $p \equiv q$ then $K * p \models \alpha$ iff $K * q \models \alpha$
R6. If $p \not\models \bot$ then $K * p \not\models \bot$
R7. If $K * (p \wedge q) \models \alpha$ then $((K * p) + q) \models \alpha$
R8. If $K * p \not\models \neg q$ then (if $((K * p) + q) \models \alpha$ then $(K * (p \wedge q)) \models \alpha$)

KM Belief Update Belief update is an approach to reasoning with changing beliefs after some fundamental shift in the world occurred. It is characterised by a belief set ψ, an update operation $\diamond$ and postulates for reasoning. As with revision, ψ refers to a logically closed set of propositional formulas. When we update ψ with new information μ, we are saying that we used to believe ψ, we know now that μ holds, and we need to modify ψ by adding μ, acknowledging that we may have been wrong if μ contradicts ψ. There are nine postulates in the KM belief update framework, described in [9] and shown in Table 2.

Table 2. KM Postulates

U1. $\psi \diamond \mu \models \mu$
U2. If $\psi \models \mu$ then $\psi \diamond \mu$ iff ψ
U3. If both ψ and μ is satisfiable then $\psi \diamond \mu$ is satisfiable
U4. If ψ_1 iff ψ_2 and μ_1 iff μ_2 then $\psi_1 \diamond \mu_1$ iff $\psi_2 \diamond \mu_2$
U5. $(\psi \diamond \mu) \wedge \phi \models \psi \diamond (\mu \wedge \phi)$
U6. If $\psi \diamond \mu_1 \models \mu_2$ and $\psi \diamond \mu_2 \models \mu_1$ then $\psi \diamond \mu_1$ iff $\psi \diamond \mu_2$
U7. If ψ is complete then $(\psi \diamond \mu_1) \wedge (\psi \diamond \mu_2) \models \psi \diamond (\mu_1 \vee \mu_2)$
U8. $(\psi_1 \vee \psi_2) \diamond \mu$ iff $(\psi_1 \diamond \mu) \vee (\psi_2 \diamond \mu)$
U9. If ψ is complete and $(\psi \diamond \mu) \wedge \phi$ is satisfiable then $\psi \diamond (\mu \wedge \phi) \models (\psi \diamond \mu) \wedge \phi$

B Methodology and Ethical Issues

Our goal is to provide empirical evidence to test our hypothesis that human reasoning is consistent with the AGM and KM postulates. Empirical evidence relies on data collected from human subjects, which typically follow three approaches: quantitative, qualitative or mixed [4]. We have chosen to follow a mixed-methods approach in our research design because our research incorporates elements of both quantitative and qualitative approaches. Many different terms are used in the literature for the mixed-methods approach, such as integrating, synthesis, quantitative and qualitative methods, multimethod and mixed methodology, but recent writings tend to use the term mixed-methods [2, 10, 12, 13]. We obtained ethical clearance from the Faculty of Science Ethics Research Committee at the University of Cape Town. Our application included information about the researcher, the research focus, participant procedures and how consent is obtained. In our research, we intend to collect information from the same number of individuals on both the qualitative and quantitative fronts. In other words, each participant will give identical numbers of quantitative and qualitative responses. No responses will be assigned particular weights. Rather, in our estimation of sample size, we will consider factors such as percentage occurrence of a state or condition (population variability), percentage maximum error and level of confidence to provide an adequate count. Typically, mixed methods researchers would include the sample of qualitative participants in the larger quantitative sample, because ultimately researchers make a comparison between the two databases and the more they are similar, the better the comparison [4]. For our research, each participant will provide both qualitative and quantitative data as both are pertinent to answering our research question. Together, the belief and explanation given by the participant will create a stronger understanding of their beliefs than either would give separately. It has been argued by Grootswagers [8] that online experiment data are generally better when experiments are short, pay well, are fun, and have clear instructions. We anticipate that each of the three surveys would take a general MTurk worker 30 to 45 minutes to complete, including reading time for the consent form and instructions. Following best practice [8, 14, 16], we will divide each of our surveys into multiple smaller HITs, also mitigating participant fatigue. Bentley et al. [3] presented a study investigating the reliability of fast survey platforms such as Amazon MTurk and SurveyMonkey as compared to larger market research studies for technology behaviour research. They found that results can be obtained in hours for much smaller costs with accuracy within 10% of traditional larger surveys [3]. Grootswagers [8] describes an overview of conducting behavioural research online. In their work, they cite online testing as advantageous as compared to traditional testing for three reasons: (i) efficiency, (ii) more representative population and (iii) more economical [8]. Our design incorporates the use of a keyboard and mouse or trackpad for user responses, allowing for efficient online testing as alluded to by Grootswagers [8].

C Experiments

In Table 3, we show the evaluation survey participants used in Experiment 1. We show visual representations of our quantitative and qualitative collected data

Table 3. Evaluation table for general reasoning statements

No.	Question	Yes	No	Comment
(a)	I acknowledge that I am a native English speaker / that I have a good understanding and command of the English language.			
(b)	I acknowledge that I have read statements 1 - 30 in full.			
(c)	Do you believe the language used in any of the statements 1 - 30 to be unclear? If you answered "Yes", please write the statement number(s) in the "Comment" column and give a reason for each.			
(d)	Do you believe the content of any of the statements 1 - 30 to be biased or implausible? If you answered "Yes", please write the statement number(s) in the "Comment" column and give a reason for each.			
(e)	Are there any statements regarding behaviours or traits of people in a certain profession which you would like to see included in this survey? If you answered "Yes", please write your suggestions under the "Comment" column.			
(f)	Any feedback to add?			

from Experiment 2 in Figures 1–2. We show the contingency tables for the AGM postulates, as part of Experiment 3, in Tables 4–12.

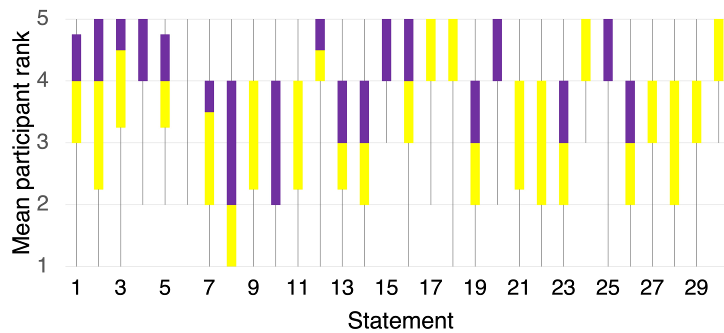

Fig. 1. Box plot of average participant rank against statement number

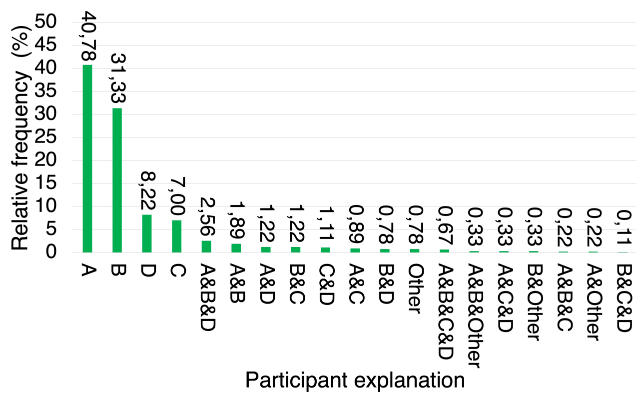

Fig. 2. Explanation category against relative frequency (%)

Table 4. R1 (i)

31	4
0	0

Table 5. R1 (ii)

34	1
0	0

Table 6. R2

31	1
2	1

We show the contingency tables for the KM postulates, as part of Experiment 4, in Tables 13–21.

Table 7. R3

26	2
7	0

Table 8. R4 (i)

30	5
0	0

Table 9. R5

4	19
5	17

Table 10. R6

32	1
0	2

Table 11. R7

31	4
0	0

Table 12. R8

28	1
4	2

Table 13. U1

29	8
0	0

Table 14. U2

25	1
4	7

Table 15. U3

31	0
4	2

Table 16. U4

13	0
7	17

Table 17. U5

35	2
0	0

Table 18. U6

24	5
0	8

Table 19. U7

21	4
7	5

Table 20. U8

20	17
0	0

Table 21. U9

25	5
4	4