Cognitive Defeasible Reasoning: the extent to which forms of defeasible reasoning correspond with human reasoning

Clayton Baker[1] [0000-0002-3157-9989], Claire Denny[1] [0000-0002-7999-8699], Paul Freund[1] [0000-0002-2826-6631], and Thomas Meyer[2] [0000-0003-2204-6969]

[1] University of Cape Town, South Africa
bkrcla003@myuct.ac.za
dnncla004@myuct.ac.za
frnpau013@myuct.ac.za

[2] University of Cape Town, South Africa and CAIR
tmeyer@cs.uct.ac.za

Abstract. Classical logic is the default for modelling human reasoning but has been found to be insufficient to do so. It lacks the flexibility so characteristically required of reasoning under uncertainty, incomplete information and new information, as people must. In response, non-classical extensions to propositional logic have been formulated, to provide non-monotonicity. We focus on three such extensions, KLM Defeasible Reasoning, AGM Belief Revision and KM Belief Update, discussing related work and reporting on three respective experiments designed to test empirically the extent to which the postulates propounded to characterise these logic forms conform to human reasoning. We find evidence to believe two properties of Defeasible Reasoning (one of which is a KLM Defeasible Reasoning property), three properties of AGM Belief Revision and four properties of KM Belief Update conform. In the cases of Defeasible Reasoning and Belief Revision, we discuss the relationship they have with human reasoning. We find evidence that suggests, overall, Defeasible Reasoning has a normative relationship and Belief Revision a descriptive relationship. In the case of Belief Update, we discuss counter-examples to the KM postulates, as tested in the experiment pertaining to it.

Keywords: propositional logic · cognitive defeasible reasoning · defeasible reasoning · belief revision · belief update · non-monotonicity · Google Forms · Mechanical Turk.

1 Introduction

Reasoning is an integral part of human lives. It is well-documented that human reasoning fails to conform to the prescriptions of classical or propositional logic [22], modelling a flexibility considered key to intelligence [20]. The artificial intelligence community therefore seeks to incorporate such flexibility in their work

[20]. Non-classical or non-monotonic logic is flexible by nature. Classical reasoning is sufficient to describe systems with a calculated output in an efficient way. However, the way in which humans reason is non-classical, as humans are known to reason in different ways [22]. The problem is that non-monotonic reasoning schemes have been developed for and tested on computers, but whether there exists a correspondence with human reasoning that needs to be investigated. This problem is important because we can gain insight into how humans reason and incorporate this into building improved non-monotonic AI systems. An issue which needs to be considered is that humans are diverse subjects - some reason normatively while others reason descriptively. In the case of normative reasoning, a reasoner would conclude that a certain condition should be the case or that the condition is usually the case. In the case of descriptive reasoning, a reasoner would make a bold claim that a certain condition is exactly true of exactly false. A comprehensive investigation needs to be done to detect the extent of the correspondence between non-monotonic reasoning and the way in which humans reason. In Section 2, we describe related work and three specific forms of non-monotonic reasoning. In Section 3, we describe our research question and additional investigation questions. In Section 4, we describe the design and implementation of three distinct surveys, each of which incorporates questions of a distinct formalism of non-monotonic reasoning and each of which seeks to determine the extent of correspondence between that formalism and human reasoning. In Section 5, we describe the methods used to analyse our survey results. We present our results, discussion and conclusions in Section 6. Lastly, we describe the track for future work in Section 7.

2 Background

Humans are known to reason differently about situations in everyday life. Non-monotonic reasoning is the study of those ways of inferring additional information from given information that do not satisfy the monotonicity property which is satisfied by all methods based on classical logic [13]. With non-monotonic reasoning, a conclusion drawn about a particular situation does not always hold. This type of reasoning is described in the context of AI [21]. We consider non-monotonic reasoning forms *Defeasible Reasoning*, *Belief Revision* and *Belief Update*. The latter two are both forms of *belief change* [11], wherein there exists a belief base and a belief set [5]. Explicit knowledge the agent has about the world resides in the base and inferences or knowledge derived from that in the base resides in the belief set.

2.1 Defeasible Reasoning

In philosophy, when a conclusion has the potential to be withdrawn, or when a conclusion can be reinforced with additional information, the conclusion is said to be defeasible. Defeasible reasoning occurs when the evidence available to the reasoner does not guarantee the truth of the conclusion being drawn [20].

A defeasible statement has two identifiable parts: an antecedent and a consequence [6]. Often, the meaning attached to a consequence, given an antecedent, is not straightforward. We shall now illustrate this with an example. Consider the following statements: *employees pay tax* and *Alice is an employee*. From the statements given, can we conclude that *Alice pays tax*? Using defeasible reasoning, we can infer that *Alice pays tax* and *Alice does not pay tax*. The conclusion *Alice pays tax* depends on whether Alice is a typical employee or whether Alice is an exceptional employee and thus does not pay tax.

2.2 Belief Revision

In belief revision, conflicting information indicates flawed prior knowledge on the part of the agent, forcing the retraction of conclusions drawn from it [11,17]. Information is then taken into account by selecting the models of the new information closest to the models of the base, where a model of information μ is a state of the world in which μ is true [11]. An example of this reasoning pattern will now be described. Consider the same statements used above in the defeasible reasoning example. Using the reasoning pattern of belief revision, we can infer from our beliefs that Alice does pay tax. Suppose we now receive new information: *Alice does not pay tax*. This is inconsistent with our belief base, so a decision must be made regarding which beliefs to retract prior to adding the new information into our beliefs. We could revise our beliefs to be that *employees pay tax* and *Alice does not pay tax*. In [4], this decision is proposed to be influenced by whether we believe some statements more strongly than others. In [1], it is proposed to be influenced by closeness (the concept of minimal change), in that we aim to change as little about our existing knowledge as we can do without having conflicting beliefs.

2.3 Belief Update

In belief update, conflicting information is seen as reflecting the fact that the world has changed (without the agent being wrong about the past state of the world). To get an intuitive grasp of the distinction between belief update and revision, take the following example adapted from [11]. Let b be the proposition that the book is on the table, and m be the proposition that the magazine is on the table. Say that our belief set includes $(b \wedge \neg m) \vee (\neg b \wedge m)$, that is the book is on the table or the magazine is on the table, but not both. We send a student in to report on the state of the book. She comes back and tells us that the book is on the table, that is b. Under the AGM postulates for belief revision proposed in [1], we would be warranted in concluding that $b \wedge \neg m$, that is, the book is on the table and the magazine is not. But consider if we had instead asked her to ensure that the book was on the table. After reporting, we again are faced with the new knowledge that b. This time adding the new knowledge corresponds to the case of belief update. And here it seems presumptuous to conclude that the magazine is not on the table [11]. Either the book was already on the table and the magazine was not, in which case the student would have done nothing and left, or the magazine was on the table and the book not, in which case the student presumably would have simply put the book on the table and left the magazine similarly so. As these examples are formally identical, there need be different formalisms to accommodate both cases.

3 Problem Statement

Research Question: To what extent do the theoretical models of defeasible reasoning, belief revision and belief update correspond with human reasoning?

We have investigated three approaches to non-monotonic reasoning: the KLM [13] defeasible reasoning approach, the AGM [1] belief revision approach and the KM [10] belief update approach. In additional investigations, the reasoning style of participants - normative or descriptive - was identified in the cases of defeasible reasoning and belief revision. For belief update, the additional investigation was to find counter-examples to the KM [10] properties.

4 Implementation

In this section, we describe the design and implementation of three surveys: one each for defeasible reasoning, belief revision, and belief update. We also describe our implementation strategy and expected challenges. We proceed to document our testing and evaluation strategy.

4.1 Survey Designs

Our survey designs refer to concrete and abstract questions. An example of a concrete, story-style or real-world question would be: *If Cathy has a cake to bake, will she use an oven?*. An example of an abstract question would be: given the following, *If A then B*, and *If C then A*, can we say that *If C then B*?.

Survey 1 The survey included 18 questions mainly pertaining to the KLM [13] properties of defeasible reasoning. The 18 questions were divided into 2 categories, namely concrete questions and abstract questions. There were 10 concrete questions and 8 abstract questions. Each question contained a reasoning scenario based on a particular property of defeasible reasoning. The conversions from formal properties to human reasoning scenarios were guided by the usual translations between English and the logics in question. Of the 10 concrete questions, 2 questions related to prototypical and presumptive reasoning respectively. The questions about prototypical and presumptive reasoning did not have corresponding abstract questions as they are not governed by any postulates. These questions instead were included because of relevance to defeasible reasoning as a whole. Questions 2a to 9a were strictly based on the concrete form of the KLM [13] properties, whilst questions 10a to 17a were based on the abstract form of those properties. The participant was then required, for each question, to state whether they agreed or disagreed with the conclusion and provide a reason for their answer. The reason for the answer given by the respondent was crucial to our study. We have used the participants' reasons to identify the reasoning style used as either normative or descriptive.

Survey 2 The questions were developed to test whether properties of a specific formalisation of the process of Belief Revision feature in cognitive reasoning. The formalisation used is that of the eight-postulate approach as proposed by Alchourrón, Gärdenfors and Makinson (AGM) [1]. Two types of questions were developed: concrete and abstract. This involved designing scenarios in which to ground the concrete questions. Five such scenarios were designed, with ten concrete questions being formulated overall. Eight abstract questions were developed, directly based on the formal properties. The abstract questions were included so as to test the properties without having the agent's knowledge of the world hindering their answers and to have questions which are less semantically loaded [16] than real-world concrete questions. The benefit of abstract examples is further discussed by Pelletier and Elio [20]. The concrete questions started out as abstract representations explicitly requiring the application of one or some of the formal properties to obtain the desired answer. These representations were then elaborated in the context of a scenario. The scenarios designed are: linguists, smoking, wildlife, bag of stationery and, acrobats. The scenarios designed are partly inspired by the literature and partly by the researcher's knowledge of the world and creativity. The linguist scenario is inspired by the former and the other four scenarios are inspired by the latter.

Survey 3 The questions were developed to test the KM approach [11] to belief update. The questions are broken into three sets. The first consists of abstract questions, in which the KM postulates are presented and people are to rate their agreement with the postulates on a linear or Likert scale with extremal points 'strongly agree' and 'strongly disagree'. The postulates are presented using non-technical language. The second is concrete questions that are meant to be confirming instances of each of the eight KM postulates, where participants were asked to answer either yes or no, and provide a reason for their answer. The third set follows the same format as the second (i.e. also contains concrete questions), however it is meant to present counter-examples to the postulates, with the counter-examples largely sourced from the literature. The first counter-example is based on the observation that updating p by $p \vee q$ has no effect in the KM approach [7], which seems counter-intuitive. The second is based on the observation that updating by an inclusive disjunction leads to the exclusive disjunction being believed in the right conditions (a modification of the checkerboard example in [7]), which again seems counter-intuitive. The third is based on the observation that sometimes belief revision semantics seem appropriate in cases corresponding to the way that belief update is commonly, and has been here, presented [15]. The final is an example testing a counter-intuitive result of treating equivalent sentences as leading to equivalent updates.

4.2 Mechanical Turk

Mechanical Turk (MTurk) is a service provided by Amazon that serves as an interface between *service requesters* and a network of humans. It addresses three problems [9]. It is used by software developers to incorporate human intelligence into software applications. It is used by business people to access a large network of human intelligence to complete tasks, for example to conduct market research. It is used by people looking to earn money to find work that can be done anywhere and at any time, using the skills they already have.
We used Mechanical Turk for access to its network of humans to complete our surveys, which were hosted on Google Forms. The advantage of Mechanical Turk is that its network of Workers includes people from many places, with a large range of ages and education levels [23]. Such places include the United States of America, Canada, India, Pakistan, the United Kingdom and the Philippines [23]. The reward per response should have a lower bound of the minimum wage, for ethical reasons [24]. The location of respondents requested for this project, however, was not restricted. This meant that the minimum wage in South Africa (R20/hour [18]) could not be used as the lower bound, as there are other countries with greater amounts as their minimum. With research feasibility in mind, in terms of funding, a compromise was made: to ensure the compensation offered was greater than the South African minimum wage, and rely on the comparatively low amount to deter those for whom MTurk is their primary source of income from accepting the task.

4.3 Google Forms

Google Forms is a survey administration application that allows users to create free online surveys. Research performed in 2018 revealed that the recent surge of low-quality qualitative data from MTurk is primarily due to international Turkers (workers on MTurk) [26] using Virtual Private Networks (VPNs) or Virtual Private Servers (VPSs) to waive qualifications required to complete surveys [12]. This motivated including a checkpoint within the surveys themselves. Given that the survey is answered online and not in person, the sections included a checkpoint, comprising of custom captchas and an attention check, designed to be an indicator of the suitability of the respondent to take the survey. In this context, suitability comprises four requirements: *(i)* the response is not generated by a bot, *(ii)* the respondent is not using a script, *(iii)* the respondent can understand English, *(iv)* the respondent reads questions in full.

4.4 Testing And Evaluation

The survey, once constructed, was tested. This entailed evaluation of the survey by a group of laypeople and experts and the publication of the survey in a HIT on MTurk as a trial. The results of the trial HIT were evaluated, to gauge both how Turkers might respond to the final survey.

Feedback from groups of laypeople and experts We asked a variety of experts and non-experts to evaluate our survey for coherence, clarity and other desirable characteristics of questions, more examples of which can be found in [14]. One of the authors evaluated each of the three surveys. We also approached an expert in Psychology and Cognitive Science and an expert in Philosophy, at the University of Cape Town, but unfortunately, they were not available to evaluate our survey. The remaining experts who evaluated the survey questions included (for particular surveys) one Master's student in computer science, as well as two computer science Honours students also conducting a study on belief revision and belief update respectively. For a general perspective, one of the surveys was evaluated by an international doctoral student in language and African studies. Based on the suggestions from experts and laypeople, a variety of changes were made to the surveys.

Trial HITS A trial of the surveys using MTurk was conducted. This was done *(i)* to gain familiarity with the MTurk service and platform and *(ii)* to test the survey and its questions on a sample of Turkers. It involved three separate postings of the survey links as HITs on the site, two of which required 5 responses and one which required 6. The HIT was created with certain specifications accordingly. Workers were compensated R30 (above the South African hourly minimum wage) for completing the tasks, and the tasks included an initial time estimate, all of which were under an hour. We did not restrict workers by location, but required that they should have a certain number of completed HITs previously, and have a certain approval rating for their tasks. Based on the results from the trial survey, various changes were made for the final experiments, some of which included increasing compensation, and changing the estimated times of completion. The process we engaged in can thus be seen as engaging in an initial prototype of the experiment and then iteratively improving that prototype.

4.5 Ethical, Professional and Legal Issues

As this project involved experiments with people, before proceeding with the experiments ethical clearance was obtained from the University of Cape Town Faculty of Science Human Research Ethics Committee. The major ethical issue in the experiments was the use of Mechanical Turk, and in particular whether workers were being paid a fair wage for their work. Per [2] the following three steps were taken to mitigate these concerns. First, workers were paid more than the South African minimum wage for an hour's work. Second, in the title of the task the estimated amount of time needed for the task was clearly stated. Finally, there is a section in the survey which gives an overview of what the research is concerned with, placing the work in context. Additionally, workers were required to give their informed consent to participate in the study. This was achieved by having a consent form at the start of the survey, whereby workers could either agree to participate in the research and then continue to the rest of the survey, or they could decline to participate and be thanked for their time. Contact details of the researchers were also provided. Before the data-handling, all survey responses were anonymised. To this end, we also did not collect names, cellphone numbers or email addresses from our participants. The only personal contact information we collected from each participant was their Amazon Turk Worker ID. Furthermore, there were no apparent professional issues as this was purely a research project.

5 Methods of Analysis

5.1 Survey 1

Assumptions and Data Overview In our collected data, the *Yes* and *No* responses were considered quantitative data whilst the reasons given by participants were considered qualitative data. We assumed that the property of *Ref* holds for all human reasoners and therefore it was not tested in our study. Upon receiving feedback from our supervisor, it was found that question 6a and 7a were not appropriate models of the properties they intended to test - *RW* and *And* respectively - because the conclusions in those questions contained the condition *some* which could not be converted to a propositional setting and thus could not be converted to their respective KLM [13] properties. The responses to questions 6a and 7a were removed from our final data set and removed from our analysis further in this study. As a result, our data set contained responses from 29 participants to 16 questions only. This gave a total of 464 closed responses and 464 leading responses.

Response Categories We identified four main emerging themes: *Support*, *Speculative*, *Technical* and *Other*. After identifying these four themes, we have qualified a normative response as being either *Support* or *Technical*. We have qualified a descriptive response as only being Speculative. The remaining responses, those classified as *Other*, were not analysed further.

Analysis One of our research questions involves determining which of the KLM [13] properties were true with human reasoning. The baseline for our analysis is calculated at the ceiling of 50% of the total number of responses obtained. The results obtained were considered significant if responses exceeded the baseline amount. A positive hit rate for a particular property indicates that the number of *Yes* responses exceeded the baseline whilst the number of *No* responses were below the baseline. In contrast, a negative hit rate for a particular property indicates that the number of *No* responses exceeded the baseline whilst the number of *Yes* responses were below the baseline.

5.2 Survey 2

Analysis of the Questions section of the survey, for both the trial and final survey, comprised finding the modal answer and hit rate for each closed question and performing qualitative analysis on the open questions. The data was downloaded from Google Forms and Mechanical Turk.

Quantitative Data The modal answer and hit rate (%) for closed questions were determined by applying functions in Microsoft Excel to the data. A hit indicates success. In this context, success is defined as both the respondent and the application of the properties of belief revision obtaining the same answer for a question. Hit rate is thus calculated for each question as $\frac{number of successes}{no. of responses} \times 100$.

Qualitative Data The qualitative analysis was performed in NVivo, a qualitative data analysis software package, and made use of *Tesch's Eight Steps in the Coding Process* [3]. In this process, a combination of pre-determined and emerging codes were used. Codes on topics expected to be found were taken from literature, based on the theory being empirically tested. These include the eight properties of belief revision as proposed by Alchourrón, Gärdenfors and Makinson [1]: closure, success, inclusion, vacuity, consistency, extensionality, super-expansion, sub-expansion. Other pre-determined codes include: normative and descriptive. Emerging codes are those which were not anticipated at the beginning, or are both unusual and of interest. They are developed solely on the basis of the data collected from respondents by means of the survey. An example of an emerging code used in the trial of this study is *It is stated*. This code represents the respondent taking a passive approach to their response. Other examples would be *real-world influence* and *likelihood*.
Pre-determined codes *normative* and *descriptive* refer to the reasoning style identified in responses to open questions. A normative style involves making value judgements [19], commenting on whether something is the way it should be or not. This includes implied judgements through the use of emotive language. A descriptive style, in contrast, does not - it involves making an observation, commenting on how something is [19].

5.3 Survey 3

Quantitative Data For the quantitative data, two forms of analysis were chosen, corresponding to the two different forms of quantitative data (ordinal and binary) gathered. For the ordinal (Likert-type) data, the median is an appropriate measure of central tendency [25], and thus was chosen, and for the binary data the hit-rate as above was chosen. Relating this back to the research question, a postulate was seen as confirmed if it saw both a hit rate of over 50% for the confirming concrete example, and a median value of agree or better.

Qualitative Data For the qualitative data, emerging codes were developed for section 2 on a *per question* basis. This was so as to better interpret the quantitative results, and, in particular with the counter-examples, see whether the reasons given by participants for their answers matched the theory behind the objections as given in the literature. Similar to the belief revision case, a common code was *new information should be believed*, which corresponds to the case of simply believing new information. The predetermined codes of normative/descriptive were not applied in this case.

6 Results, Discussion and Conclusions

To answer our research question, we found several correspondences between the KLM [13] approach for defeasible reasoning, the AGM [1] approach for belief revision and the KM [10] approach for belief update, and the way in which our participants reasoned. For defeasible reasoning, there was correspondence in four instances: correspondence with the KLM [13] defeasible reasoning property of *Or*, correspondence with the defeasible reasoning property of *Transitivity* as well as correspondence with two types of defeasible reasoning: *prototypical* and *presumptive* reasoning. For belief revision, there was correspondence with three AGM [1] properties: *Success*, *Vacuity* and *Closure*. For belief update, there was correspondence with four KM [10] properties: *U1*, *U2*, *U4* and *U6*. For each of the three surveys, we present additional results that are of importance.

6.1 Additional results for Defeasible Reasoning

For the concrete and abstract KLM [13] property of *RM*, the hit rates had a significant discrepancy of >50%, whilst the properties of *LLE* and *CM* had discrepancies of >30% between hit rates. It was thus inconclusive as to whether there was a positive or negative relationship between human reasoning and these defeasible reasoning properties. For the concrete and abstract properties of *And* and *RW*, responses could not be used due to the questions not adhering to the propositional nature of the properties, through the use of the keyword *some*. Further research needs to be done to establish the relationship between human reasoning and the defeasible properties of *And* and *RW* in concrete form. Through an additional investigation, we also found that participants have a normative relationship with defeasible reasoning when properties are presented both in concrete and abstract form. However, participants reasoned less normatively when faced with abstract reasoning situations as compared to concrete reasoning situations.

6.2 Additional results for Belief Revision

The hit rates were taken as indications of the type of relationship between human reasoning and the relevant AGM [1] properties. For the concrete and abstract questions for properties *Success*, *Closure* and *Vacuity*, the hit rates obtained were $\geq 50\%$, suggesting a positive relationship between human reasoning and those properties. Properties *Extensionality*, *Super-expansion* and *Consistency* received hit rates $\leq 50\%$, suggesting a negative relationship. Properties *Sub-expansion* and *Inclusion* had discrepancies of $> 30\%$ between the hit rates for their concrete and abstract questions, and their relationships to human reasoning thus found to be inconclusive. Through an additional investigation, we found that participants have a predominantly descriptive relationship with belief revision when properties are presented both in concrete and abstract form. The balance of descriptive and normative reasoning styles of respondents in their responses became more even for the abstract questions, perhaps suggesting an increasing reliance on perceived rules in situations to which humans are less able to relate.

6.3 Additional results for Belief Update

Every confirming concrete example of the properties saw a hit rate $> 50\%$, and only *U7* and *U8* saw a neutral or worse median value for the abstract questions. Additionally, the codes from the confirming instances of the KM postulates revealed an interesting theme. For all postulates excluding *U5* and *U7*, the codes for the majority reason for agreement with the examples were closely related to the postulates in question. For example, for *U1* the majority code was *new information should be believed.* The results thus provide some indication that the KM postulates model a more general trend in reasoning, if we accept that reasons are based on general themes. All of the counter-examples examined saw hit rates $>50\%$. The first counter-example followed from *U2*, which states that updating on something already logically implied has no effect. The second counter-example (modification of Herzig and Rifi's [8] checkerboard example) follows from the set of *U1*, *U4*, and *U5*. The third counter-example was Lang's [15] against *U8*, the hallmark of belief update which semantically differentiates it from belief revision. The final counter-example follows independently from *U4*, that syntax is irrelevant to update, and the set of *U1* and *U6*. Although *U1* is implicated theoretically, the codes developed for the responses here indicate that participants who agreed with the counter-examples believed the new information in them, hence the results did not support that *U1* is violated.

7 Future work

While the three forms of non-monotonic reasoning examined are meant to be a better model of human reasoning than propositional logic, the results of this project indicate that they are not yet a perfect fit, with participants failing to reason in accordance with many of the properties of the systems. Future work involving conducting a study with a larger participant pool is necessary to obtain more accurate results. It may also be interesting to add blocks, in the form of different control groups, to the study to explore the effects of different circumstances on cognitive reasoning and which logic form is most closely resembled in each such block. Further avenues include a more direct comparison of survey results.

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. Buhrmester, M.: M-turk guide (2018), `https://michaelbuhrmester.wordpress.com/mechanical-turk-guide/`
3. Creswell, J.W.: Research design: qualitative, quantitative, and mixed methods approaches, pp. 245–253. SAGE Publications, Thousand Oaks, California, 4 edn. (2014)
4. Darwiche, A., Pearl, J.: On the logic of iterated belief revision. Artificial Intelligence **89**, 1–29 (01 1997). https://doi.org/10.1016/S0004-3702(96)00038-0
5. Gärdenfors, P.: Belief Revision: An Introduction, pp. 1–26. Cambridge University Press (01 1992). https://doi.org/10.1017/CBO9780511526664.001
6. Guido, G., Terenziani, P.: Temporal extensions to defeasible logic. Australasian Joint Conference of Artificial Intelligence pp. 476–485 (2007)
7. Herzig, A., Rifi, O.: Update operations: a review. In: Prade, H. (ed.) Proceedings of the 13th European Conference on Artificial Intelligence. pp. 13–17. John Wiley and Sons (1998)
8. Herzig, A., Rifi, O.: Propositional belief base update and minimal change. Artificial Intelligence **115**(1), 107–138 (1999)
9. Inc., A.M.T.: Faqs (2018), `https://www.mturk.com/help`
10. Katsuno, H., Mendelzon, A.O.: On the difference between updating a knowledge base and revising it. In: Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning. pp. 387–394. KR'91, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1991), `http://dl.acm.org/citation.cfm?id=3087158.3087197`
11. Katsuno, H., Mendelzon, A.O.: On the difference between updating a knowledge base and revising it. Belief Revision **29**, 183 (2003)
12. Kennedy, R., Clifford, S., Burleigh, T., Jewell, R., Waggoner, P.: The shape of and solutions to the mturk quality crisis (10 2018)
13. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artif. Intell. **44**, 167–207 (1990)
14. Krosnich, J., Presser, S.: Question and questionnaire design. Handbook of Survey Research (03 2009)
15. Lang, J.: Belief update revisited. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence. pp. 1534–1540,2517–2522. IJCAI'07, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2007), `http://dl.acm.org/citation.cfm?id=1625275.1625681`
16. Lehmann, D.: Another perspective on default reasoning. Annals of Mathematics and Artificial Intelligence **15**(1), 61–82 (1995)
17. Martins, J., Shapiro, S.: A model for belief revision. Artif. Intell. **35**, 25–79 (01 1988). https://doi.org/10.1016/0004-3702(88)90031-8
18. Omarjee, L.: Everything you need to know about the national minimum wage (01 2019), `https://www.fin24.com/Economy/everything-you-need-to-know-about-the-national-minimum-wage-20190101`
19. Over, D.: Rationality and the normative/descriptive distinction. In: Koehler, D.J., Harvey, N. (eds.) Blackwell handbook of judgment and decision making. pp. 3–18. Blackwell Publishing Ltd, United States (2004)

20. Pelletier, F., Elio, R.: The case for psychologism in default and inheritance reasoning. Synthese **146**, 7–35 (08 2005). https://doi.org/10.1007/s11229-005-9063-z
21. Pollock, J.: A theory of defeasible reasoning. International Journal of Intelligent Systems **6**, 33–54 (1991)
22. Ragni, M., Eichhorn, C., Kern-Isberner, G.: Simulating human inferences in light of new information: A formal analysis. In: Kambhampati, S. (ed.) Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI 16). pp. 2604–2610. IJCAI Press (2016)
23. Ross, J., Zaldivar, A., Irani, L., Tomlinson, B.: Who are the turkers? worker demographics in amazon mechanical turk (01 2009)
24. Semuels, A.: The internet is enabling a new kind of poorly paid hell (01 2018), `https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/`
25. Sullivan, G., A.R. Artino, J.: Analyzing and interpreting data from likert-type scales. Journal of Graduate Medical Education **5**(4), 541–542 (2013)
26. TurkPrime: After the bot scare: Understanding what's been happening with data collection on mturk and how to stop it (09 2018), `https://blog.turkprime.com/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it`

8 SUPPLEMENTARY INFORMATION

8.1 Defeasible Reasoning

KLM Properties Table 1 presents the KLM postulates. We use $\alpha \mid\!\sim \gamma$ to represent that a statement, α, defeasibly entails a statement, γ.

Table 1. KLM Postulates

1. Reflexivity	$\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \alpha$
2. Left Logical Equivalence	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma,\ \mathcal{K} \not\models\!\!\approx \alpha \mid\!\sim \neg\beta}{\mathcal{K} \models\!\!\approx \alpha \wedge \beta \mid\!\sim \gamma}$
3. Right Weakening	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \rightarrow \beta, \gamma \mid\!\sim \alpha}{\mathcal{K} \models\!\!\approx \gamma \mid\!\sim \beta}$
4. And	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \beta, \mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma}{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \beta \wedge \gamma}$
5. Or	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma, \mathcal{K} \models\!\!\approx \beta \mid\!\sim \gamma}{\mathcal{K} \models\!\!\approx \alpha \vee \beta \mid\!\sim \gamma}$
6. Cautious Monotonicity	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \beta, \mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma}{\mathcal{K} \models\!\!\approx \alpha \wedge \beta \mid\!\sim \gamma}$

Additional Properties Table 2 presents additional defeasible reasoning postulates.

Table 2. Additional Postulates

1. Cut	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \wedge \beta \mid\!\sim \gamma,\ \mathcal{K} \models\!\!\approx \alpha \mid\!\sim \beta}{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma}$
2. Rational Monotonicity	$\dfrac{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma,\ \mathcal{K} \not\models\!\!\approx \alpha \mid\!\sim \neg\beta}{\mathcal{K} \models\!\!\approx \alpha \wedge \beta \mid\!\sim \gamma}$
3. Transitivity	$\dfrac{\alpha \mid\!\sim \beta,\ \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma}$
4. Contraposition	$\dfrac{\alpha \mid\!\sim \beta}{\neg\beta \mid\!\sim \neg\alpha}$

Survey Questions For access to the survey questions on Google Forms, click here.

Survey Data For access to the raw and coded data, click here.

8.2 Belief Revision

Properties Table 3 presents the AGM postulates. $K * \alpha$ is the sentence representing the knowledge base after revising the knowledge base K with α.

Table 3. AGM Postulates

1. Closure	$\mathrm{K} * \alpha = C_n(\mathrm{K} * \alpha)$
2. Success	$\mathrm{K} * \alpha \models \alpha$
3. Inclusion	$\mathrm{K} * \alpha \subseteq C_n(\mathrm{K} \vee \{\alpha\})$
4. Vacuity	If $\neg\alpha \notin \mathrm{K}$ then $C_n(\mathrm{K} \vee \{\alpha\}) \subseteq \mathrm{K} * \alpha$
5. Consistency	$\mathrm{K} * \alpha = C_n(\alpha \wedge \neg\alpha)$ only if $\models \neg\alpha$
6. Extensionality	If $\alpha \equiv \phi$ then $\mathrm{K} * \alpha = \mathrm{K} * \phi$
7. Super-expansion	$\mathrm{K} * (\alpha \wedge \phi) \subseteq C_n(\mathrm{K} * \alpha \vee \{\phi\})$
8. Sub-expansion	If $\neg\phi \notin \mathrm{K}$ then $C_n(\mathrm{K} * \alpha \vee \{\phi\}) \subseteq \mathrm{K} * (\alpha \wedge \phi)$

Survey Questions For access to the survey questions on Google Forms, click here. A clean copy of the final survey will be shown, with the required fields for the checkpoint, main body of questions and feedback questions relaxed for ease of viewing.

Survey Data For access to the raw data and quantitative analysis, click here. For access to the NVivo file containing the coded data, click here. For access to the code book used in the qualitative analysis, click here.

8.3 Belief Update

Properties Table 4 presents the KM postulates. $\phi \diamond \alpha$ is the sentence representing the knowledge base after updating the knowledge base represented by ϕ with α.

Table 4. KM Postulates

(U1) $\phi \diamond \alpha \models \alpha$
(U2) If $\phi \models \alpha$ then $\phi \diamond \alpha = \phi$
(U3) If both ϕ and α are satisfiable then $\phi \diamond \alpha$ is satisfiable
(U4) If $\phi_1 \leftrightarrow \phi_2$ and $\alpha_1 \leftrightarrow \alpha_2$ then $\phi_1 \diamond \alpha_1 \leftrightarrow \phi_2 \diamond \alpha_2$
(U5) $(\phi \diamond \alpha) \wedge \gamma \models \phi \diamond (\alpha \wedge \gamma)$
(U6) If $\phi \diamond \alpha_1 \models \alpha_2$ and $\phi \diamond \alpha_2 \models \alpha_1$ then $\phi \diamond \alpha_1 \leftrightarrow \phi \diamond \alpha_2$
(U7) If ϕ is complete then $(\phi \diamond \alpha_1) \wedge (\phi \diamond \alpha_2) \models \phi \diamond (\alpha_1 \vee \alpha_2)$
(U8) $(\phi_1 \vee \phi_2) \diamond \alpha \leftrightarrow (\phi_1 \diamond \alpha) \vee (\phi_2 \diamond \alpha)$

Survey Questions For access to the survey questions on Google Forms, click here.

Survey Data For access to the raw data, click here. For access to coded data and some qualitative analysis, click here.

8.4 More information

For information beyond what is provided here, view our project website.