# 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

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. Non-monotonic reasoning refers to making an inference which is not absolute: in light of new information, the inference could change. We focus on three extensions of non-monotonic reasoning, KLM Defeasible Reasoning [6], AGM Belief Revision [2] and KM Belief Update [5]. We have investigated, via surveys, the extent to which each of KLM Defeasible Reasoning, AGM Belief Revision and KM Belief Update correspond with human 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 In philosophy, when a conclusion has the potential to be withdrawn, or when a conclusion can be reinforced with additional information, the conclusion is said to be defeasible.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 [6] [9]. For Defeasible Reasoning, we investigated the KLM properties of *Left Logical Equivalence*, *Right Weakening*, *And*, *Or* and *Cautious Monotonicity*. We find evidence for correspondence with the KLM property of *Or*, which states that any formula that is, separately, a plausible consequence of two different formulas, should also be a plausible consequence of their disjunction. We also investigate conformance with human reasoning and two subtypes of Defeasible Reasoning: prototypical [7] and presumptive reasoning [11]. Prototypical reasoning is an approach that suggests each reasoning scenario assumes a prototype with certain typical features, whereas presumptive reasoning suggests that an argument may have multiple possible consequences. We find that both subtypes of defeasible reasoning conform. In Belief Revision, conflicting information indicates flawed prior knowledge on the part of the agent, forcing the retraction of conclusions drawn from it [5,8]. 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 $\mu$ is a state of the world in which $\mu$ is true [5]. For Belief Revision, we investigated the AGM properties of *Clo-*

*sure*, *Success*, *Inclusion*, *Vacuity*, *Consistency*, *Extensionality*, *Super-expansion* and *Sub-expansion*. We find evidence for correspondence with the AGM property of *Success*, which expresses that the new information should always be part of the new belief set. We also find evidence for correspondence with the AGM properties of *Closure* and *Vacuity*. *Closure* implies logical omniscience on the part of the ideal reasoner, including revision of their belief set. *Vacuity* is motivated by the principle of minimal change and suggests that if the incoming sentence is not in the original set, then there is no effect. Literature suggests a formal link between Defeasible Reasoning and Belief Revision. We take a step towards investigating whether this formal link translates to an empirical link. Thus, 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. A normative [3] [10] relationship suggests that humans reason according to believed norms accepted in general by other human reasoners, whereas a descriptive [1] [4] relationship indicates that humans choose to consider external sources of information as additional grounds on which to make an inference. 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. For Belief Update, we investigated the KM postulates *U1*, *U2*, *U3*, *U4*, *U5*, *U6*, *U7* and *U8*, as seen in Table 4 of the Appendix. We find evidence for correspondence with postulate *U1*, which states that updating with the new fact must ensure that the new fact is a consequence of the update. We find evidence for correspondence with postulate *U3*, which states the reasonable requirement that we cannot lapse into impossibility unless we either start with it, or are directly confronted by it. We also find evidence for correspondence with postulates U4 and *U6*. Postulate *U4* asserts that syntax is irrelevant to the results of an update. Postulate *U6* states that if updating on $\alpha_1$ entails $\alpha_2$ and if updating on $\alpha_2$ entails $\alpha_1$, then the effect of updating on either is equivalent. In the literature, the KM postulates for Belief Update have seen less acceptance than the AGM postulates for Belief Revision. To this end, we discuss counter-examples to the KM postulates, as tested in the experiment pertaining to it. 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. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI communications **7**(1), 39–59 (1994). https://doi.org/10.3233/AIC-1994-7104
2. 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
3. Artosi, A., Cattabriga, P., Governatori, G.: An automated approach to normative reasoning pp. 132–145 (1994)
4. Besold, T.R., Uckelman, S.L.: Normative and descriptive rationality: from nature to artifice and back. Journal of Experimental & Theoretical Artificial Intelligence **30**(2), 331–344 (2018). https://doi.org/10.1080/0952813X.2018.1430860, `https://doi.org/10.1080/0952813X.2018.1430860`
5. Katsuno, H., Mendelzon, A.O.: On the difference between updating a knowledge base and revising it. Belief Revision **29**, 183 (2003)
6. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence **44**, 167–207 (1990)
7. Lieto, A., Minieri, A., Piana, A., Radicioni, D.: A knowledge-based system for prototypical reasoning. Connection Science **27**(2), 137–152 (2015). https://doi.org/10.1080/09540091.2014.956292
8. 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
9. Pelletier, F., Elio, R.: The case for psychologism in default and inheritance reasoning. Synthese **146**(2), 7–35 (2005)
10. Torre, L.V.D., Tan, Y.: Diagnosis and decision making in normative reasoning. Artificial Intelligence and Law **7**(1), 51–67 (1999). https://doi.org/10.1023/A:1008359312576
11. Verheij, B.: Correct grounded reasoning with presumptive arguments. European Conference on Logics in Artificial Intelligence pp. 481–496 (2016). https://doi.org/10.1007/978-3-319-48758-8_31


# 1 SUPPLEMENTARY INFORMATION

## 1.1 External resources

We have created a GitHub repository which contains additional resources. In this repository, we include our survey questions, the coding of the survey responses as well as our complete project paper. The GitHub repository can be accessed by clicking here. In addition, a summary of our project work is also showcased on our project website which can be viewed by clicking here.

## 1.2 Defeasible Reasoning

**KLM Properties** Table 1 presents the KLM postulates. We use $\alpha \mid\!\sim \gamma$ to represent that a statement, $\alpha$, defeasibly entails a statement, $\gamma$. Reflexivity states that if a formula is satisfied, it follows that the formula can

**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 | $\dfrac{\mathcal{K} \models\!\!\approx \alpha \mid\!\sim \beta, \mathcal{K} \models\!\!\approx \alpha \mid\!\sim \gamma}{\mathcal{K} \models\!\!\approx \alpha \wedge \beta \mid\!\sim \gamma}$ |

be a consequence of itself. Left Logical Equivalence states that logically equivalent formulas have the same consequences. Right Weakening expresses the fact that one should accept as plausible consequences all that is logically implied by what one thinks are plausible consequences. And expresses the fact that the conjunction of two plausible consequences is a plausible consequence. Or says that any formula that is, separately, a plausible consequence of two different formulas, should also be a plausible consequence of their disjunction. Cautious Monotonicity expresses the fact that learning a new fact, the truth of which could have been plausibly concluded, should not invalidate previous conclusions.

**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 | $\dfrac{\alpha \mid\!\sim \beta}{\neg\beta \mid\!\sim \neg\alpha}$ |

Cut expresses the fact that one may, in his way towards a plausible conclusion, first add an hypothesis to the facts he knows to be true and prove the plausibility of his conclusion from this enlarged set of facts and then deduce (plausibly) this added hypothesis from the facts. Rational Monotonicity expresses the fact that only additional information, the negation of which was expected, should force us to withdraw plausible conclusions previously drawn. Transitivity expresses that if the second fact is a plausible consequence of the first and the third fact is a plausible consequence of the second, then the third fact is also a plausible

consequence of the first fact. Contraposition allows the converse of the original proposition to be inferred, by the negation of terms and changing their order.

### 1.3 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 $\alpha$.

**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 $C_n(\mathrm{K} * \alpha \vee \{\phi\}) \subseteq \mathrm{K} * (\alpha \wedge \phi)$ |

Closure implies logical omniscience on the part of the ideal agent or reasoner, including after revision of their belief set. Success expresses that the new information should always be part of the new belief set. Inclusion and Vacuity is motivated by the principle of minimum change. Together, they express that in the case of information $\alpha$, consistent with belief set or knowledge base K, belief revision involves performing expansion on K by $\alpha$ i.e. none of the original beliefs need to be withdrawn. Consistency expresses that the agent should prioritise consistency, where the only acceptable case of not doing so is if the new information, $\alpha$, is inherently inconsistent - in which case, success overrules consistency. Extensionality effectively expresses that the content i.e. the belief represented, and not the syntax, affects the revision process, in that logically equivalent sentences or beliefs will cause logically equivalent changes to the belief set. Super-expansion and sub-expansion is motivated by the principle of minimal change. Together, they express that for two propositions $\alpha$ and $\phi$, if in revising belief set K by $\alpha$ one obtains belief set K' consistent with $\phi$, then to obtain the effect of revising K with $\alpha \wedge \phi$, simply perform expansion on K' with $\phi$. In short, $K * (\alpha \wedge \phi) = (K * \alpha) + \phi$.

### 1.4 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 $\phi$ with $\alpha$.

U1 states that updating with the new fact must ensure that the new fact is a consequence of the update. U2 states that updating on a fact that could in principle be already known has no effect. U3 states the reasonable requirement

**Table 4.** KM Postulates

| |
|---|
| (U1) $\phi \diamond \alpha \models \alpha$ |
| (U2) If $\phi \models \alpha$ then $\phi \diamond \alpha = \phi$ |
| (U3) If both $\phi$ and $\alpha$ 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 $\phi$ is complete then $(\phi \diamond \alpha_1) \wedge (\phi \diamond \alpha_2) \models \phi \diamond (\alpha_1 \vee \alpha_2)$ |
| (U8) $(\phi_1 \vee \phi_2) \diamond \alpha \leftrightarrow (\phi_1 \diamond \alpha) \vee (\phi_2 \diamond \alpha)$ |

that we cannot lapse into impossibility unless we either start with it, or are directly confronted by it. U4 requires that syntax is irrelevant to the results of an update. U5 says that first updating on $\alpha$ then simply adding the new information $\gamma$ is at least as strong (i.e. entails) as updating on the conjunction of $\alpha$ and $\gamma$. U6 states that if updating on $\alpha_1$ entails $\alpha_2$ and if updating on $\alpha_2$ entails $\alpha_1$, then the effect of updating on either is equivalent. U7 applies only to complete $\phi$, that is $\phi$ which have only one model. If some situation arises from updating a complete $\phi$ on $\alpha_1$ and it also results from updating that $\phi$ from $\alpha_2$ then it must also arise from updating that $\phi$ on $\alpha_1 \vee \alpha_2$ [5]. U8 is the disjunction rule.

### 1.5 Results

In Figure 1, we show the Hit Rate (%) for each Defeasible Reasoning Postulate. In Figure 2, we show the Hit Rate (%) for Prototypical Reasoning and Presumptive Reasoning. In Figure 3, we show the Hit Rate (%) for each Belief Revision Postulate. In Figure 4, we show the Hit Rate (%) for each Belief Update Postulate.

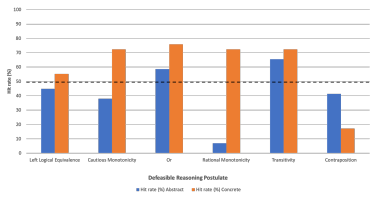


**Fig. 1.** Hit Rate (%) for each Defeasible Reasoning Postulate

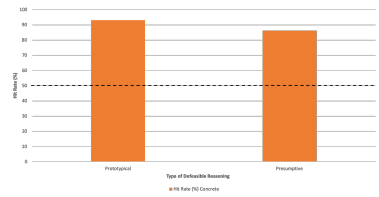


**Fig. 2.** Hit Rate (%) for Prototypical Reasoning and Presumptive Reasoning

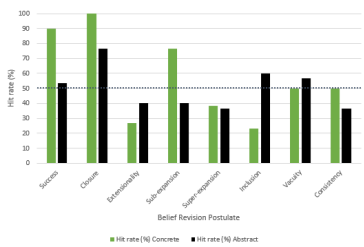


**Fig. 3.** Hit Rate (%) for Belief Revision Postulates

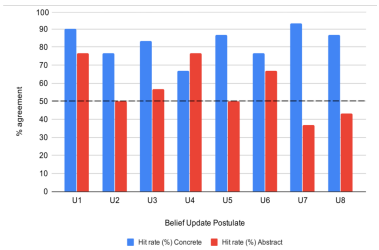


**Fig. 4.** Hit Rate (%) for Belief Update Postulates