How Defeasible and Nonmonotonic Reasoning relates to Human Reasoning

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

May 1, 2019

Abstract

In this paper, we explore four types of reasoning. We look at classical reasoning as a starting point and then describe three nonmonotonic reasoning approaches. We will look at formal descriptions of defeasible reasoning, belief revision and belief update. We aim to conclude that there is a need for further investigation into the closeness of these types of nonmonotonic reasoning, and how humans reason.

Keywords

defeasible reasoning, nonmonotonic reasoning

1 Introduction

In classical reasoning, a conclusion follows deductively from a premise. A premise, also called an antecedent, is a set of given information from which a reasoner can draw a conclusion. Classical reasoning is monotonic because once the conclusion is warranted on the basis of certain premises, no amount of additional information will invalidate the conclusion. In everyday life, conditional statements such as "if Lee-Anne has to bake a cake, then she will use an oven" are used to describe, for example, rules, regularities or causal chains [3]. A conditional statement consists of an antecedent, and a consequence which holds in the classical logical sense, if the antecedent is satisfied. It is well known from behavioural studies that human reasoning is complex and does not generally follow formal classical reasoning logic. Humans do not always endorse a consequence, even if the antecedent is satisfied. In addition, humans do not always deduce from a negated consequence that the antecedent is false. This paper aims to go beyond classical reasoning as it is clear that humans do not, in general, reason monotonically.

Nonmonotonic 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 [1]. Defeasible reasoning occurs when the evidence available to the reasoner does not guarantee the truth of the conclusion being drawn [2]. We will now describe an example. From the statements, "Linguists typically speak more than three languages" and "Kim is a linguist," one might infer, by default, "Kim speaks more than three languages". The meaning of "default" is that we are justified in making this inference because we have no information which would make us doubt that Kim was a typical linguist or that Kim was an exception to the the statement about linguists in this regard [2].

With belief revision, additional information which is inconsistent with our old beliefs, makes our current beliefs inconsistent with the new information. The original antecedent has to be revised in order to remain consistent [4]. In a different way, belief update looks at a change not in the current beliefs, but a change in the world itself. This change in the world does not allow us to conclude that some of the old worlds were not possible and a new model of the world has to be evaluated [5]. As a result, this paper is interested in the extent to which human reasoning corresponds to these types of reasoning. In particular, this paper aims to identify the extent to which human reasoning resonates with defeasible reasoning and therefore this literature review will focus on defeasible reasoning.

2 Background

In this section, we provide a formal description of the operators and rules for classical reasoning. We shall use a language, L, which is based on propositional logic. It will be assumed that L is closed under applications of the boolean connectives $\neg$ (negation), $\wedge$ conjunction, $\vee$ disjunction, and $\rightarrow$ implication. We will use Greek letters such as α, β and γ etc. as variables over sentences in L. We also introduce the tautology constants $\top$ and $\bot$ as "truth" and "falsity" respectively. All the different beliefs of the world expectations will be formulated in L. We will assume that the underlying logic includes classical propositional logic and that it is compact. Recall that a logic is said to be compact iff whenever α is a logical consequence of a set A of sentences, then there is a finite subset A' of A such that α is a logical consequence of A'.

If A logically entails α, we will write this as $A \vdash \alpha$. In classical entailment, $\vdash$ satisfies the deduction theorem and "disjunction in the premises" i.e. that $A \cup \{\beta \vee \gamma\} \vdash \alpha$ whenever both $A \cup \{\beta\} \vdash \alpha$ and $A \cup \{\gamma\} \vdash \alpha$ [6]. Other details of L are left open. Where A is a set of sentences we shall use the notation $C_n(A) = \{\alpha : A \vdash \alpha\}$. The nonmonotonic inference relation will be denoted by $\mid\!\sim$. We will also introduce the notation $C(\alpha)$ for the set of all nonmonotonic conclusions that can be drawn from α, that is, $\beta \in C(\alpha)$ iff $\alpha \mid\!\sim \beta$.

3 Defeasible Reasoning

Defeasible reasoning is a way of inferring a consequence from an antecedent which is normally true. If α and β are formulas, then the pair $\alpha \mid\!\sim \beta$ (read *if α, normally β* or *β is a plausible consequence of α*) is called a conditional assertion [1]. The antecedent of the assertion is the formula α and β is its consequent. The meaning we attach to such an assertion is the following: if α is true, we are willing to infer that normally β is true [1]. From this, we can see that there is a need to investigate the meaning of 'normally', not only as an operator, but as understood by humans. In the following section, we will discuss sets of conditional assertions called *consequence relations*.

3.1 Practical implementation

We shall now motivate that the study of nonmonotonic consequence relations will be a benefit to the field of human reasoning. The queries one wants to ask an automated knowledge base are formulas (of L) and the query β should be interpreted as: *is β expected to be true?* To give an answer, the knowledge base will use some approach to draw a conclusion from the information that it has. Similarly, humans wish to answer a query from a situation that may arise, such as "if it is sunny outside, can I expect that it will be windy?". In the following, we suggest a description of the different types of information an automated knowledge base could have and then propose that humans have similar distinctions between different types of knowledge that they have.

The first type of information we describe is the beliefs of the world, U, that describe both hard constraints (e.g. fish are vertebrates) and points of definition (e.g. youngster is not equivalent to adult) [1]. Equivalently, such information can be described by a set of formulas defining U to be the set of all worlds that satisfy all the formulas of this set.

The second type describes soft constraints (e.g. people normally drink milk) using a set of conditional assertions [1]. This set describes our beliefs about the general behaviour of the world. The set of conditional assertions will be called the knowledge base, and denoted by $\mathcal{K}$.

The third type of information describes our information about the particular situation at hand (e.g. it is a rhino). We represent this information by a formula. Furthermore, an alternative to regarding the first type of information as a separate type, is given by Kraus, Lehmann and Magidor [1] as follows: "One could, equivalently, treat a formula α of the first type as the conditional assertion $\neg\alpha \mid\!\sim$ **false**. One could also have decided to introduce all information of the third type as information of the first type".

We now describe the way in which our inference procedure will work, to answer query β. Given the context of the universe of reference, U, it will try to deduce (in a way yet unknown) the conditional assertion $\alpha \mid\!\sim \beta$ from the knowledge base $\mathcal{K}$. From the above, it is clear that any system of nonmonotonic reasoning may be considered as such. However, we would like to investigate the extent to which humans deduce conditional assertions from the knowledge that they have.

3.2 Semantics

We shall now describe the meaning of conditional assertions. Let α, β and γ be formulas. The conditional assertion A is $\gamma \wedge \alpha \mid\!\sim \beta$. The conditional assertion B is $\gamma \mid\!\sim \neg\alpha \vee \beta$, i.e. $\gamma \mid\!\sim \alpha \rightarrow \beta$. We shall only consider the casewhen the formula γ is a tautology. In this case, A is $\alpha \mid\!\sim \beta$ and B is **true** $\mid\!\sim \neg\alpha \vee \beta$. The assertion A says that *if α, normally β*. Assertion B says that *Normally, if α is true then β is true*. A and B have very different meanings, notably when α is normally false. Assertion A says that in this exceptional case when α is true, one also expects β to be true. Assertion B, then, becomes true if α is normally false. From this, it implies that is is likely that B does not say anything about cases when α is true (if these cases are exceptions).

Take for example α to be "it is an ostrich" and β to be "it flies". In the context of birds, it seems logical to accept B which says that *normally, either it is not an os-*

trich or it flies, since normally birds fly. Nevertheless, one should be reluctant to accept A which says ostriches normally fly. It becomes clear that A and B have two different meaning and that A does not follow from B [1].

3.3 Properties

The following properties for defeasible reasoning were formalised by Kraus, Lehmann and Magidor [1]]: .

1. Reflexivity (REF)

$\mathcal{K} \approx \alpha \mid\sim \alpha$

Reflexivity states that if a formula is satisfied, it follows that the formula can be a consequence of itself.

2. Left Logical Equivalence (LLE)

$$\frac{\beta \equiv \alpha, \mathcal{K} \approx \alpha \mid\sim \gamma}{\mathcal{K} \approx \beta \mid\sim \gamma}$$

Left Logical Equivalence states that logically equivalent formulas have the same consequences.

3. Right Weakening (RW)

$$\frac{\mathcal{K} \approx \alpha \mid\sim \beta, \beta \models \gamma}{\mathcal{K} \approx \alpha \mid\sim \gamma}$$

Right Weakening expresses the fact that one should accept as plausible consequences all that is logically implied by what one thinks are plausible consequences.

4. And

$$\frac{\mathcal{K} \approx \alpha \mid\sim \beta, \mathcal{K} \approx \alpha \mid\sim \gamma}{\mathcal{K} \approx \alpha \mid\sim \beta \wedge \gamma}$$

And expresses the fact that the conjunction of two plausible consequences is a plausible consequence.

5. Or

$$\frac{\mathcal{K} \approx \alpha \mid\sim \gamma, \mathcal{K} \approx \beta \mid\sim \gamma}{\mathcal{K} \approx \alpha \vee \beta \mid\sim \gamma}$$

Or says that any formula that is, separately, a plausible consequence of two different formulas, should also be a plausible consequence of their disjunction.

6. Cautious Monotonicity (CM)

$$\frac{\mathcal{K} \approx \alpha \mid\sim \beta, \mathcal{K} \approx \alpha \mid\sim \gamma}{\mathcal{K} \approx \alpha \wedge \beta \mid\sim \gamma}$$

Cautious Monotonicity expresses the fact that learning a new fact, the truth of which could have been plausibly concluded, should not invalidate previous conclusions.

7. Rational Monotonicity (RM)

$$\frac{\mathcal{K} \approx \alpha \mid\sim \gamma,\ \mathcal{K} \not\approx \alpha \mid\sim \neg\beta}{\mathcal{K} \approx \alpha \wedge \beta \mid\sim \gamma}$$

Rational Monotonicity expresses the fact that only additional information, the negation of which was expected, should force us to withdraw plausible conclusions previously drawn.

4 Related work

In this section we will describe belief revision and belief update as nonmonotonic reasoning approaches related to defeasible reasoning.

4.1 Belief Revision

Belief revision is a way of representing dynamic knowledge. It is how we modify our beliefs when we receive new information. A problem arises when that information is inconsistent with the beliefs that represent our current beliefs. For instance, suppose we believe that a Mercedes Benz is the fastest car and then we found out that some BMW cars are faster than any Mercedes Benz cars. As a result, we must revise our beliefs in order to accept the new information while maintaining as much of the old information as possible.

One of the most controversial properties of the revision operators is success. Success specifies that the new information is stronger than the beliefs of a reasoner. In [7], a semi-revision operator is also introduced. The semi-revision operator sometimes accepts the new information and sometimes rejects it. Furthermore, if the new information is accepted, then deletions from the old information are made if this is necessary to maintain consistency. There is also a close connection between the semi-revision and merge operators (detail not discussed in this paper, but can be found in [7]) and the frameworks for default reasoning. The idea is that the beliefs to be deleted in a knowledge base could be preserved in an alternate set as defeasible rules or assumptions.

Alchourrón, Gärdenfors and Makinson [4] have suggested eight postulates which they argue must be satisfied by any reasonable revision function. These postulates are formulated in a very general setting, but we restrict the discussion here to the propositional logic case. Instead of a finite knowledge base, they consider a *knowledge set*, that is, a set of formulas that is deductively closed. Given a knowledge set $\mathcal{K}$ and sentence μ, $\mathcal{K} * \mu$ is the revision of $\mathcal{K}$ by μ. The postulates for belief revision from [4] are given as follows:

1. $\mathcal{K} * \alpha = C_n(\mathcal{K} * \alpha)$

2. $\mathcal{K} * \mu \models \mu$

3. $\mathcal{K} * \mu \subseteq C_n(\mathcal{K} \cup \{\mu\})$

4. $\dfrac{\neg\mu \notin \mathcal{K}}{C_n(\mathcal{K} \cup \{\mu\}) \leq \mathcal{K} * \mu}$

5. $\mathcal{K} * \mu = C_n(\mu \wedge \neg\mu) \leftrightarrow \models \neg\mu$

6. $\dfrac{\mu \equiv \phi}{\mathcal{K} * \mu = \mathcal{K} * \phi}$

7. $\mathcal{K} * (\mu \vee \phi) \leq C_n(\mathcal{K} * \mu \cup \{\phi\}$

8. $\dfrac{\neg\phi \notin \mathcal{K} * \mu}{C_n(\mathcal{K} * \mu \cup \{\phi\}) \subseteq \mathcal{K} * (\mu \wedge \phi)}$

Alchourrón, Gärdenfors and Makinson [4] also suggest that if there is a way of representing any knowledge set $\mathcal{K}$ by a propositional formula ψ such that $\mathcal{K} = \{\phi | \psi \models \phi\}$, then we can establish a direct link between $\mathcal{K} * \mu$ and $\psi \circ \mu$.

4.2 Belief Update

Belief update consists of bringing the knowledge base up to date when the world described by it changes. A typical database entry could be of the form, e.g. ”increase Amy’s annual salary by 2.5%”. Suppose knowledge base ψ is to be updated with sentence μ. Update methods will select, for each model M of the knowledge base ψ, the set of models μ that are closest to M.

Suppose that ψ has exactly two models, I and J; that is, there are two possible worlds described by the knowledge base. Suppose μ describes exactly two worlds, K and L, and that K is ”closer” to I than L is, and K is also closer to I than L is to J. Then K is selected for the new knowledge base, but L is not. Note the knowledge base has effectively forgotten that J used to be a possible world; the new fact μ has been used as evidence for the retroactive impossibility of J. That is, not only do we refuse to have J as a model of the new knowledge base, but we also conclude that J should not have been in the old knowledge base to begin with [5]. The models of ψ are possible worlds; we think one of them is the real world, but we do not know which one. Now the real world has changed; we examine each of the old possible worlds and find the minimal way of changing each one of them so that it becomes a model of μ. The fact that the real world has changed gives us no grounds to conclude that some of the old worlds were actually not possible [5].

Katsuno and Mendelzon [5] proposed eight postulates which they argue must be observed for any belief update function. We use $\psi \diamond \mu$ to denote the result of updating $\mathcal{K}$ with μ. The postulates from [5] for belief update are given as follows:

1. $\psi \diamond \mu \models \mu$

2. $\dfrac{\psi \models \mu}{\psi \diamond \mu \equiv \psi}$

3. $\dfrac{\not\models \neg\psi,\ \not\models \neg\mu}{\not\models \neg(\psi \diamond \mu)}$

4. $\dfrac{\psi_1 \equiv \psi_2, \mu_1 \equiv \mu_2}{\psi_1 \diamond \mu_1 \equiv \psi_2 \diamond \mu_2}$

5. $(\psi \diamond \mu) \wedge \phi \models \psi \diamond (\mu \wedge \phi)$

6. $\dfrac{\psi \diamond \mu_1 \models \mu_2, \psi \diamond \mu_2 \models \mu_1}{\psi \diamond \mu_1 \equiv \psi \diamond \mu 2}$

7. $\dfrac{\text{If } \psi \text{ is complete}}{(\psi \diamond \mu_1) \wedge (\psi \diamond \mu_2) \models \psi \diamond (\mu_1 \vee \mu_2))}$

8. $(\psi_1 \vee \psi_2) \diamond \mu \equiv (\psi_1 \diamond \mu) \cup (\psi_2 \diamond \mu)$

5 Discussion

For each of the four models discussed in this paper, the literature focusses on the meaning of the operators involved and ability of the reasoner to be flawed. In the next example, we will show that humans do reason nonmonotonically. Suppose a reasoner was presented with the following information ”oranges are sweet” and ”Liam bought an orange”, then he would likely conclude that ”Liam’s orange is sweet”. In this case, if Liam tasted his orange and revealed that it was, in fact, sour, the reasoner is forced to retract his original conclusion. This behaviour corresponds with the description of nonmonotonic reasoning in this paper. We would like to further investigate to what extent nonmonotonic and, in particular, defeasible reasoning relates to human reasoning.

6 Conclusion

Classical reasoning can be described as a linear form of reasoning. It consists of statements, contained in a knowledge base, and classical operators which can be used to draw conclusions. With classical reasoning, a conclusion β is classically entailed by a premise α with notation, $\alpha \vdash \beta$. β is thus a logical consequence of α. Once the conclusion is justified by the reasoner, using the classical approach, no amount of additional information will make the original conclusion untrue.

In contrast, defeasible reasoning is a way of inferring a plausible consequence from an antecedent that is normally true. That is, if the antecedent α is true, then the reasoner is willing to jump to the defeasible conclusion that β is true. We say that α defeasibly entails β with notation $\alpha \mid\!\sim \beta$.

We looked at belief revision as another form of non-monotonic reasoning which is a way of representing dynamic knowledge. When a reasoner is presented with additional information which is inconsistent with his old beliefs, the inconsistent beliefs have to be removed from the old beliefs.

Lastly, we have looked at belief update. Belief update is a way of updating a knowledge base when the world described by it changes. In this approach, the reasoner will select the models of the world, contained in the knowledge base, that are closest to the knowledge of the new world.

These four approaches have been studied in the literature as formal theoretical models and we would like to investigate the closeness of the connection between these forms of reasoning and human reasoning. We also propose an investigation into the particular relation between human reasoning and defeasible reasoning.

7 References

[1] Sarit Kraus, Daniel Lehmann and Menachem Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence 44 (Jul. 1990), 167-207. DOI: https://doi.org/10.10 16/0004-3702(90)90101-5

[2] Francis J. Pelletier and Renee Elio. 2005. The Case for Psychologism in Default and Inheritance Reasoning. Synthese 146 (2005), 7-35. DOI: https://doi.org/10.1007/s11229-005-9063-z

[3] Marco Ragni, Christian Eichhorn, Tanja Bock, Gabriele Kern-Isberner and Alice Tse. 2017. Formal Nonmonotonic Theories and Properties of Human Defeasible Reasoning. Minds and Machines 27 (2017), 79-117. DOI: https://doi.org/10.1007/s11023-016-9414-1

[4] Hirofumi Katsuno and Alberto O. Mendelzon. 1991. Propositional knowledge base revision and minimal change. Artificial Intelligence 52, 3 (Dec. 1991), 263-294. DOI: https://doi.org/10.1016/0004-3702(91) 90069-V

[5] Hirofumi Katsuno and Alberto O. Mendelzon. 2003. On the Difference between Updating a Knowledge Base and Revising it. Belief revision 29 (2003), 183.

[6] Peter Gardenfors and David Makinson. 1994. Nonmonotonic inference based on expectations. Artificial Intelligence 65, 2 (1994), 197-245. DOI: https://doi.org/10.10 16/0004-3702(94)90017-5

[7] Marcelo A. Falappa, Gabriele Kern -Isberner and Guillermo R. Simari. 2002. Explanations, belief revision and defeasible reasoning. Artificial Intelligence 141 (2002), 1-28. DOI: https://doi.org/10.10 16/S0004-3702(02)00258-8

[8] Daniel Lehmann. 1998. Stereotypical reasoning: Logical Properties. Logic journal of the IGPL 6, 1 (1998), 49-58. DOI: https://doi.org/10.1093/jigpal/6.1.49

[9] Daniel Lehmann. 1995. Another perspective on default reasoning. Artificial Intelligence 15, 1 (1995), 61 - 85. DOI: https://doi.org/10.1007/BF01535841

[10] John L. Pollock. 1991. A Theory of Defeasible Reasoning. International Journal of Intelligent Systems 6, 1 (Jan. 1991), 33-54. DOI: https://doi.org/10.1002/ int.4550060103

[11] Daniel Lehmann and Menachem Magidor. 1992. What does a conditional knowl-

edge base entail? Artificial Intelligence 55, 1 (1992), 1-60. DOI: https://doi.org/10.1016/0004-3702(92)90041-U

[12] John L. Pollock. 1987. Defeasible Reasoning. Cognitive Science 11, 4 (1987), 481-518. DOI: https://doi.org/10.1016/S0364-0213(87)80017-4

[13] Craig Boutillier and Moises Goldszmidt. Revision by Conditional Beliefs. AAAI (Jul. 1993), 649-654.

[14] Jérôme Lang. 2014. Actions, belief update, and DDL. Krister Segerberg on Logic of Actions. Springer, Dordrecht (2014), 229-251.

[15] Johan van Benthem. 2007. Dynamic logic for belief revision. Journal of Applied Non-classical Logics 17, 2 (2007), 129-155. DOI: https://doi.org/10.3166/jancl.17.129-155

[16] Hirofumi Katsuno and Alberto O. Mendelzon. 1989. A Unified View of Propositional Knowledge Base Updates. Artificial Intelligence 11, 2 (1989), 1413-1419.