



Semantics of Belief Change Operators for Intelligent Agents

Iteration, Postulates, and Realizability

Kai Sauerwald¹

Received: 21 June 2023 / Accepted: 15 December 2023
© The Author(s) 2024

Abstract

This paper summarises several contributions to the theory of belief change by the authors' dissertation thesis. First, a relational characterization of belief revision for Tarskian logics is considered, encompassing first-order predicate logic, description logic, modal logics and many monotonic logics with model-theoretic semantics. Those logics where total preorders are the standard semantics for revision are characterized. The second contribution considered is a theory of belief revision that builds upon the idea that agents are limited in what the outcome of a revision is. Furthermore, advancements in principles for iterated belief contraction given in the thesis are outlined.

Keywords Iterated belief change · Tarskian logics · Revision · Contraction

Mathematics Subject Classification 03B42 · 68T27 · 68T30

1 Introduction

The ability to change their beliefs is a central characteristic of intelligent agents. In Knowledge Representation and Reasoning, the subarea of *belief change* [8] investigates fundamental principles, characterizations, and interrelations of principles for belief changes. The following two kinds of belief changes are of central interest:

Revision. The process of incorporating new beliefs, while maintaining consistency, whenever possible.

Contraction. The process of revoking beliefs without introducing new beliefs.

In this paper, we will consider several contributions to the theories of revision and contraction given in the authors' dissertation thesis [18]. Due to space reasons, we will focus on the high-level ideas and descriptions and refer for the full details to the corresponding references [7, 18, 19].

2 Brief Introduction to the Theory of Belief Change

We start with visiting several aspects that are considered in the theory of belief change:

Agents' Capability to Deduce and Minimal Change. A goal of belief change is to take the agent's ability to make deductions into account. If an agent is convinced that A implies B ($A \rightarrow B$) and that A holds, then she also believes in B (and other implications). Because of that, to successfully achieve that the agent believes in the contrary of B ($\neg B$) without believing in inconsistencies, the agent also has to give up at least one of the beliefs A and $A \rightarrow B$. These changes should be made minimally, in the sense that as much as possible of the initial beliefs are preserved.

Agents' Preferences. When giving up beliefs, there are often multiple possible beliefs the agent could give up (to achieve success). One can understand how these choices are made as an expression of the agent's preferences (in the space of all possible choices). We consider the preceding example. Giving up the belief in B involves the choice of giving up the belief in A or giving up the belief in $A \rightarrow B$. If the agent chooses to give up A and not $A \rightarrow B$, then we can say that the agent prefers $A \rightarrow B$ over A . However, other agents might have different preferences, e.g., another agent could choose A and $A \rightarrow B$,

✉ Kai Sauerwald
kai.sauerwald@fernuni-hagen.de; kai@kai-sauerwald.de

¹ Fakultät für Mathematik und Informatik, FernUniversität in Hagen, Künstliche Intelligenz, 58084 Hagen, Germany

then we could say that for this agent A and $A \rightarrow B$ are equally preferable.

Conditionals. Conditionals and belief changes are interrelated via Ramsey-test-like connections [18, 21]: B is believed after revision by A if and only if the conditional $(B \mid A)$ is accepted by the agent. Hence, sets of conditionals provide a semantics for belief changes.

Iterated Belief Change. The way an agent adjusts her beliefs may be influenced by prior changes. We consider the previous example again for further elucidation. Assume that the agent has chosen to give up A for believing in $\neg B$, but to keep $A \rightarrow B$. The agent then might believe neither in A nor in B ($\neg A \wedge \neg B$). What would happen if the agent becomes aware after some time that A or B ($A \vee B$) has to be true? The agent has to choose how to establish belief in $A \vee B$ by starting to believe in, e.g., in A , in B or fully in $A \vee B$. However, in light of the prior change, the agent might consider some choices as more reasonable. For example, because the belief in B was just revoked, she might not consider B as a good choice.

The area of belief change employs mathematical logic. Given a logic $\mathbb{L} = (\mathcal{L}, \Omega, \vDash)$, the formulas $\alpha \in \mathcal{L}$ of the logic are employed to describe beliefs, the interpretations $\omega \in \Omega$ stand for possible worlds (that the agent can imagine), and logical entailment represents the agent’s reasoning capabilities. Belief change is the process of changing a set of formulas K by new information α to a new set $K \circ \alpha$, whereby \circ stands for any belief change operator. A prominent and widely accepted logic-based approach to revision and contraction is by Alchourrón, Gärdenfors and Makinson (AGM) [1] that captures the principle of minimal change when K and $K \circ \alpha$ are deductively closed, especially when \mathbb{L} is a propositional logic. In this paper, each of “AGM revision” and “AGM contraction” contraction refer to the respective full set of postulates, including the supplementary ones [1], or an adaption of these for the respective framework.

3 Characterization of Revision in Tarskian Logics

Agents are confronted with choosing which beliefs they want to give up when doing revisions. A very prominent approach to represent such choices is by “preferences” over possible worlds. Roughly, the connection is as follows [9]:

When revising K by α , obtain $K \star \alpha$ by selecting all those beliefs whose model set is a superset of the most “preferred” worlds where α holds. (†)

Specifically for propositional logic, for every K , the changes on K by \star are AGM revisions if and only if there is a total preorder such that one obtains $K \star \alpha$ as in (†) [9, 14]. This gives rationale to say that for propositional logic, every

AGM revision operator is *total-preorder-representable*, which we express informally as follows:

$$\text{AGM revision} \simeq \text{total preorders.} \tag{1}$$

Because preorders are considered as basic structure for representing preferences in knowledge representation [20] and beyond [15], there is theoretical interest in (1). However, for many logics it is unclear whether (1) holds.

The author’s dissertation focuses on Tarskian logics, i.e., logics where the logical consequence is a closure operator. This includes many classical monotonic logics with a model theory, e.g., first-order predicate logic, propositional logic, many modal logics, description logics, etc. Surprisingly, it turns out that (1) does not hold in some Tarskian logics; but, for Tarskian logics a correspondence similar to (1) holds [7]:

$$\text{AGM revision} \simeq \text{total min-retractive relations,} \tag{2}$$

where min-retractivity is the following weakening of transitivity [7], demanding closure under certain refinements in \leq :

For all Γ and all $\omega, \omega' \in \llbracket \alpha \rrbracket$ holds

$$\omega \in \min(\llbracket \Gamma \rrbracket, \leq) \text{ and } \omega' \leq \omega \text{ imply } \omega' \in \min(\llbracket \Gamma \rrbracket, \leq).$$

When going back to, e.g., propositional logic, the property of being total and min-retractive coincides with being a total preorder. This gives rise to the quest to identify those logics, where (1) still holds.

A main contribution given in the author’s dissertation thesis is a characterization of those Tarskian logics where every AGM revision operator is total-preorder-representable, i.e., (1) holds for every AGM revision operator. For this result, the notion of a critical loop is central, which we present in a simplified version for the sake of comprehensibility.

Definition 1 Let \mathbb{L} be a Tarskian logic. $\Gamma_{0,1}, \Gamma_{1,2}, \Gamma_{2,0}$ form a *critical loop (of length 3)* for \mathbb{L} if there exists K and consistent $\Gamma_0, \Gamma_1, \Gamma_2$ such that

- (I) $\llbracket K \cup \Gamma_{0,1} \rrbracket = \llbracket K \cup \Gamma_{1,2} \rrbracket = \llbracket K \cup \Gamma_{2,0} \rrbracket = \emptyset$,
- (II) $\llbracket \Gamma_0 \rrbracket \cup \llbracket \Gamma_1 \rrbracket \subseteq \llbracket \Gamma_{0,1} \rrbracket$,
 $\llbracket \Gamma_1 \rrbracket \cup \llbracket \Gamma_2 \rrbracket \subseteq \llbracket \Gamma_{1,2} \rrbracket$,
 $\llbracket \Gamma_2 \rrbracket \cup \llbracket \Gamma_0 \rrbracket \subseteq \llbracket \Gamma_{2,0} \rrbracket$, and
 $\llbracket \Gamma_j \cup \Gamma_i \rrbracket = \emptyset$ for each $i, j \in \{0, 1, 2\}$ with $i \neq j$, and
- (III) for each Γ_{∇} that is consistent with Γ_0, Γ_1 , and Γ_2 , there exists some consistent Γ'_{∇} such that:
 - (a) $\llbracket \Gamma'_{\nabla} \rrbracket \subseteq \llbracket \Gamma_{\nabla} \rrbracket$
 - (b) $\llbracket \Gamma'_{\nabla} \rrbracket \cap (\llbracket \Gamma_{0,1} \rrbracket \cup \llbracket \Gamma_{1,2} \rrbracket \cup \llbracket \Gamma_{2,0} \rrbracket) = \emptyset$.

Note that for a general characterization, the notion of critical loops has to be extended to arbitrary lengths [7, 18].

Given such a notion, a Tarskian logic \mathbb{L} is called *loop-free* if there is no critical loop (of any length) for \mathbb{L} . The general result can be then summarized as follows.

Theorem 2 ([7, 18]) *Let \mathbb{L} be a Tarskian logic. Every AGM revision operator for \mathbb{L} is total-preorder representable if and only if \mathbb{L} is loop-free.*

We illustrate the notion of critical loop and Theorem 2 by an example that is based on Delgrande and Peppas [6].

Example 3 We consider Horn-logic over $\Sigma = \{p_1, p_2, p_3\}$, which is a Tarskian logic. Let $\Gamma_{0,1}$, $\Gamma_{1,0}$ and $\Gamma_{2,0}$ given by

$$\begin{aligned} \Gamma_{0,1} &= \{(p_1), (p_2 \wedge p_3 \rightarrow \perp)\}, \\ \Gamma_{1,2} &= \{(p_3), (p_1 \wedge p_2 \rightarrow \perp)\}, \text{ and} \\ \Gamma_{2,0} &= \{(p_2), (p_1 \wedge p_3 \rightarrow \perp)\}. \end{aligned}$$

We show $\Gamma_{0,1}, \Gamma_{1,2}, \Gamma_{2,0}$ is a critical loop, by showing that the following K , and $\Gamma_0, \Gamma_1, \Gamma_2$ are as described in Definition 1:

$$\begin{aligned} K &= \{(p_1 \wedge p_2 \wedge p_3)\} & \Gamma_0 &= \{(p_1 \wedge p_2), (p_3 \rightarrow \perp)\} \\ \Gamma_1 &= \{(p_1 \wedge p_3), (p_2 \rightarrow \perp)\} & \Gamma_2 &= \{(p_2 \wedge p_3), (p_1 \rightarrow \perp)\} \end{aligned}$$

In the following, we list the models of these sets:

$$\begin{aligned} \llbracket K \rrbracket &= \{p_1 p_2 p_3\} & \llbracket \Gamma_0 \rrbracket &= \{p_1 p_2 \bar{p}_3\} \\ \llbracket \Gamma_1 \rrbracket &= \{p_1 \bar{p}_2 p_3\} & \llbracket \Gamma_2 \rrbracket &= \{\bar{p}_1 p_2 p_3\} \\ \llbracket \Gamma_{0,1} \rrbracket &= \{p_1 p_2 \bar{p}_3, p_1 \bar{p}_2 p_3, p_1 \bar{p}_2 \bar{p}_3\} \\ \llbracket \Gamma_{1,2} \rrbracket &= \{p_1 \bar{p}_2 p_3, \bar{p}_1 p_2 p_3, \bar{p}_1 \bar{p}_2 p_3\} \\ \llbracket \Gamma_{2,0} \rrbracket &= \{p_1 p_2 \bar{p}_3, \bar{p}_1 p_2 p_3, \bar{p}_1 p_2 \bar{p}_3\} \end{aligned}$$

One can check quickly that Condition (I) and Condition (II) are satisfied by considering the intersections of the corresponding sets of models. For Condition (III), one has to check all Γ_{∇} that are consistent with each of $\Gamma_{0,1}, \Gamma_{1,0}$, and $\Gamma_{2,0}$. In this logic, only $\Gamma_{\nabla} = \emptyset$ satisfies this condition. Because $\Gamma'_{\nabla} = K$ is consistent, and (a) and (b) from Definition 1 hold, Condition (III) is also satisfied.

Because $\Gamma_{0,1}, \Gamma_{1,2}, \Gamma_{2,0}$ is a critical loop, Theorem 2 yields that there is an AGM revision operator \star that is not total-preorder-representable. The operator \star behaves as follows:

$$\begin{aligned} \llbracket K \star (p_3 \wedge p_2 \rightarrow \perp) \rrbracket &= \{p_1 p_2 \bar{p}_3\} \\ \llbracket K \star (p_2 \wedge p_1 \rightarrow \perp) \rrbracket &= \{p_1 \bar{p}_2 p_3\} \\ \llbracket K \star (p_1 \wedge p_3 \rightarrow \perp) \rrbracket &= \{\bar{p}_1 p_2 p_3\} \end{aligned}$$

Fig. 1 illustrates the critical loop and illustrates the connection of a non-transitive part of a min-retractive relation that encodes the behaviour of \star shown above (via (\dagger)).

Because Tarskian logics with a disjunction connective are always loop-free, we obtain the following corollary,

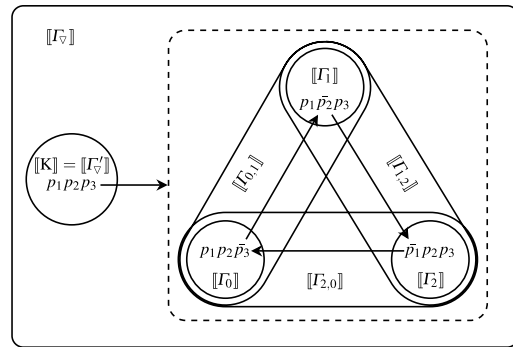


Fig. 1 Graphical illustration of a critical loop from Example 3. The areas within borders represent sets of models. The arrows depict a non-transitive relation between the interpretations; each arrow indicates that the interpretation at the start of the arrow is strictly preferred over the interpretation(s) at the arrow tip

showing that for most well-known logics, like first-order predicate logic, the connection (1) still holds.

Corollary 4 *If \mathbb{L} is a Tarskian logic that supports disjunction, then every AGM revision operator for \mathbb{L} is total-preorder representable.*

4 Revision with Limited Scope

Not every agent accepts all beliefs for revision, as, e.g., agents make distinctions based on the quality and credibility of beliefs and thus ignore certain new information. Because of this, it has been stated [10] that AGM revision operators are not an adequate model of how agents treat beliefs for revision, in the sense that agents do not accept all beliefs for revision, whereas AGM revision assumes that every belief is accepted for revision.

A prominent approach is credibility-limited revision [2, 11], which develops the idea that an agent accepts only credible new information for revision. The concept of credibility includes that an agent’s beliefs are inherently credible. Consequently, when performing credibility-limited revision, the agents want to keep as much of their current beliefs as possible during revision (as they are credible). This leads to a kind of belief revision which is guarded by a set of credibles \mathcal{C} invoking the following dynamics:

$$\llbracket K \star \alpha \rrbracket = \begin{cases} \llbracket K \rrbracket \cap \llbracket \alpha \rrbracket & \text{if } \llbracket K \rrbracket \cap \llbracket \alpha \rrbracket \neq \emptyset \\ \llbracket \alpha \rrbracket \cap \mathcal{C} & \text{if } \llbracket \alpha \rrbracket \cap \mathcal{C} \neq \emptyset \text{ and } \llbracket K \rrbracket \cap \llbracket \alpha \rrbracket = \emptyset \\ \llbracket K \rrbracket & \text{otherwise} \end{cases}$$

The dissertation thesis develops a novel theory of revision for propositional logic where the outcome of revisions is limited to certain predetermined beliefs. Developing this idea formally leads to *dynamic-limited revision*, an approach formalized in the same style as credibility-limited revision. Dynamic-limited revision involves a set of acceptables \mathcal{S} (also called the scope), and dynamic-limited revision behaves as follows:

$$\llbracket K \star \alpha \rrbracket = \begin{cases} \llbracket \alpha \rrbracket \cap \mathcal{S} \neq \emptyset & \text{if } \llbracket \alpha \rrbracket \cap \mathcal{S} \neq \emptyset \\ \llbracket K \rrbracket & \text{otherwise} \end{cases}$$

Fig. 2 provides examples of the induced belief dynamics. The long-term motivation to study dynamic-limited revision is to find good formal approaches with inherent expressiveness and explainability to describe real-world processes, e.g., transitions in psychological systems and engineering systems. Because of its different philosophical approaches, the difference to credibility-limited revision is that \mathcal{S} is not necessarily interrelated with the current beliefs of the agent. For example, K might contain outdated beliefs, yet the agent does not realise this, but as soon she revises K according to new information, she becomes aware.

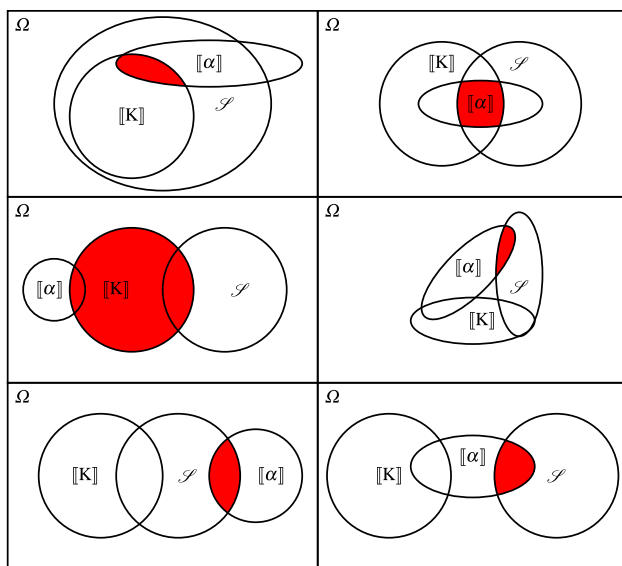


Fig. 2 Exemplary semantic description of the dynamics of dynamic-limited revision. K denotes the initial beliefs, \mathcal{S} is the set of acceptables, and α is the newly arriving information. The red area is the area containing the revision result $K \star \alpha$

5 Iteration Principles for Contraction

Iterated belief change is about how prior changes influence subsequent changes, e.g., if $K\alpha\alpha\beta$ is different from $K\alpha\gamma\beta$. A different perspective on iterated change can be obtained by employing the characterization of belief changes by preference relations: the problem of iterated belief change can be understood as the question of how to obtain a relation \leq' that characterizes $K\circ\alpha\circ\beta$ from the relation \leq that characterizes $K\circ\alpha$. Principles and approaches for the iteration of AGM revision are a well-studied topic in belief change [8, 17]. The situation is different for AGM contraction, as there is a much smaller body of literature specially devoted to the topic of principles for iterated contraction, e.g., [4, 12, 16, 17].

The author’s thesis discovers several principles for iterated contraction in propositional logic. The main source of inspiration are principles for iterated revision, e.g., by Darwiche and Pearl [5]. The following types of principles are considered for iterated contraction:

- Natural contraction.* This is the kind of change where the contraction strategy does not evolve.
- Substantiation of changes.* When β_1 and β_2 are both consistent with α , or both are inconsistent with α , then the relative credibility of β_1 and β_2 does not change when contracting α .
- Non-improvement of negation.* When contracting α the credibility of α shall not be improved.
- Independence.* In analogy to the independence principle for revision [3, 13], this means that contraction of α enforces improvement of the credibility of $\neg\alpha$.

Furthermore, gentle versions of the above-mentioned principles are considered. Besides investigating principles of change, the thesis also provides a coherent picture and a landscape of novel and known principles of iterated contraction. In particular, for all considered contraction principles, the thesis provides corresponding postulates from three different viewpoints (and proof of the equivalence of these postulates):

- the viewpoint of changing beliefs,
- the viewpoint of changing conditional beliefs, and
- the viewpoint of changing “preferences”.

Note that for some of these operations, the description of how conditional beliefs are affected is an additional novelty.

6 Summary

We considered several contributions to the theory of belief change from [18]: novel characterizations of belief revision in arbitrary Tarskain logics; dynamic-limited belief revision

operators, an approach for revision where agents are limited in what are the outcomes of a revision; advancements in the theory of iterated belief contraction by an investigation of several principles and their interrelation. Whereas all these results employ a semantic perspective on belief change.

Acknowledgements I would like to thank the reviewers for their constructive and helpful comments. The dissertation thesis was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the Priority Research Program “Intentional Forgetting in Organizations” (SPP 1921; grants BE 1700/9-1 and BE 1700/10-1 awarded to Christoph Beierle).

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alchourrón CE, Gärdenfors P, Makinson D (1985) On the logic of theory change: partial meet contraction and revision functions. *J Symb Log* 50(2):510–530
- Booth R, Fermé E, Konieczny S, Pino Pérez R (2012) Credibility-limited revision operators in propositional logic. *KR 2012*. AAAI Press, pp 116–125
- Booth R, Meyer TA (2006) Admissible and restrained revision. *J Artif Intell Res* 26:127–151
- Chopra S, Ghose A, Meyer TA, Wong K (2008) Iterated belief change and the recovery axiom. *J Philos Logic* 37(5):501–520
- Darwiche A, Pearl J (1997) On the logic of iterated belief revision. *Artif Intell* 89:1–29
- Delgrande JP, Peppas P (2015) Belief revision in Horn theories. *Artif Intell* 218:1–22
- Falakh FM, Rudolph S, Sauerwald K (2022) Semantic characterizations of AGM revision for tarskian logics. *RuleML+RR 2022, LNCS*, vol 13752. Springer, pp 95–110
- Fermé EL, Hansson SO (2018) Belief change - introduction and overview. *Springer briefs in intelligent systems*. Springer
- Grove A (1988) Two modellings for theory change. *J Philos Logic* 17(2):157–170
- Hansson SO (1999) A survey of non-prioritized belief revision. *Erkenntnis* 50(2):413–427
- Hansson SO, Fermé EL, Cantwell J, Falappa MA (2001) Credibility limited revision. *J Symb Log* 66(4):1581–1596
- Hild M, Spohn W (2008) The measurement of ranks and the laws of iterated contraction. *Artif Intell* 172(10):1195–1218
- Jin Y, Thielscher M (2007) Iterated belief revision, revised. *Artif Intell* 171(1):1–18
- Katsuno H, Mendelzon AO (1992) Propositional knowledge base revision and minimal change. *Artif Intell* 52(3):263–294
- Kelly JS (2013) *Social choice theory: An introduction*. Springer Science & Business Media
- Konieczny S, Pino Pérez R (2017) On iterated contraction: syntactic characterization, representation theorem and limitations of the Levi identity. In: Moral S, Pivert O, Sánchez D, Marín N (eds) *Scalable Uncertainty Management-11th International Conference, {SUM} 2017, Granada, Spain, Proceedings*, vol 10564. *Lecture Notes in Computer Science*. Springer, Granada, Spain, pp 348–362. https://doi.org/10.1007/978-3-319-67582-4_25
- Rott H (2009) *Shifting Priorities: Simple representations for twenty-seven iterated theory change operators*. Springer, Netherlands, pp 269–296
- Sauerwald K (2022) *Semantics of belief change operators for intelligent agents: iteration, oostulates, and realizability*, *Dissertations in Artificial Intelligence*, vol. 352. IOS Press
- Sauerwald K, Kern-Isberner G, Beierle C (2020) A conditional perspective for iterated belief contraction. *ECAI 2020*. IOS Press, pp 889–896
- Spohn W (1988) Ordinal conditional functions: a dynamic theory of epistemic states. *Causation in Decision, Belief Change, and Statistics*, vol II. Kluwer Academic Publishers, pp 105–134
- Stalnaker R (1968) A theory of conditionals. In: Rescher N (ed) *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*. Blackwell, Oxford, pp 98–112