Credibility-Limited Improvement Operators

Richard Booth[1] and **Eduardo Fermé**[2] and **Sébastien Konieczny**[3] and **Ramón Pino Pérez**[4]

Abstract. In this paper we introduce and study *credibility-limited improvement operators*. The idea is to accept the new piece of information if this information is judged credible by the agent, so in this case a revision is performed. When the new piece of information is not credible then it is not accepted (no revision is performed), but its plausibility is still improved in the epistemic state of the agent, similarly to what is done by improvement operators. We use a generalized definition of Darwiche and Pearl epistemic states, where to each epistemic state can be associated, in addition to the set of accepted formulas (beliefs), a set of credible formulas. We provide a syntactic and semantic characterization of these operators.

1 INTRODUCTION

In the logic of theory change, the AGM [1] model has acquired the status of standard model. The AGM model aims at characterizing the dynamics of beliefs of a rational agent. A change consists in adding or removing a sentence from a set of beliefs to obtain a new set of beliefs. This change obeys the following principles: 1. Primacy of new information: the new information is always accepted. 2. Coherence: the new set of beliefs has to be logically consistent. 3. Minimal change: a minimal loss of information contained in the previous beliefs, i.e., it attempts at retaining as much of the old beliefs as possible.

Even though the AGM model is considered as a standard model, it is not adequate in all contexts. Consequently, in the last 30 years extensions and generalizations of AGM have been proposed [10].

Among these extensions, we can mention: **1. Non-prioritized belief revision:** The AGM model always accepts the new information (success condition). This feature can appear unrealistic in some contexts, since rational agents, when confronted with information that strongly conflicts with their current beliefs, often reject it altogether or accept only parts of it. In non-prioritized revision, the success postulate is relaxed by weaker conditions that do not accept the new information in certain cases. **2. Iteration:** A drawback of the AGM definition of revision is that the conditions for the iteration of the process are very weak, and this is caused by the lack of expressive power of logical belief sets. In order to ensure good properties for the iteration of the revision process, a more complex structure is needed. So shifting from logical belief sets to epistemic states and additional postulates were proposed in [9].

The following example shows a case where these two extensions are needed:

Example 1 *Assume that you are an amateur painter and just finished a painting. There exists the possibility of submitting it to a contest. However, you do not believe that your painting is good. Later, you show successively the painting to different people, and all of them agree that you made a good job. At the beginning you still believe that your painting is not good enough, but, at some point in the process, you come to accept the possibility of your painting being good enough. From this moment on, it is enough for you to receive another positive judgement of your painting in order to change your mind completely and believe that your painting is good.*

Among the extensions proposed in the literature we are interested particularly in two:

1. Credibility-Limited Revision: This is based on the assumption that some inputs are accepted, others not. Those that are potentially accepted constitute the set C of credible sentences. If α is credible, then α is accepted in the revision process, otherwise no change is made to the belief set. This model was proposed and characterized for a single revision step in [11] and extended to cover iterated revision in [5].

2. Improvement operators: These operators do not (necessarily) satisfy the success postulate, although still improving the plausibility of the new information [15, 14]. This idea is quite intuitive since usual iterated belief revision operators can be considered as too strong: after revising by a new information, this information will be believed. Most of the time this is the desired behavior for the revision operators. But in some cases it may be sensible to take into account the new information more cautiously. Maybe because we have some confidence in the source of the new information, but not enough to accept it unconditionally. This can be seen as a kind of learning/reinforcement process: each time the agent receives a new information (from independent sources), this formula will gain in plausibility in the epistemic state of the agent. And if the agent receives the same new information many times, then he will finally believe it.

Credibility-limited revision operators are a quite natural formalization of non-prioritized revision, with a set of credible formulas that encodes which changes the agent can directly accept or not. But when the new information is not credible, it is simply rejected, so it does not change anything in the epistemic state. This can be seen as too drastic a position. Suppose that an agent receives many evidences from reliable (and independent) sources that a non-credible information is true. As inconceivable as this information could be for him at the beginning, the agent will surely finally reconsider its credibility.

So in this paper we propose to define a formal model of the previously described situation. This is done via the *credibility-limited improvement* operators where, when the new information is not credible, the agent performs an improvement. So the plausibility of this information increases, and sufficiently many iterations can finally lead to acceptance of this information.

The structure of the paper is as follows: we begin with a section of preliminary concepts. Section 3 is devoted to the rationality postulates. Section 4 contains the main result of the paper: a representation theorem[5]. In Section 5 we give some examples in order to illustrate

[1] University of Luxembourg, Luxembourg, E-mail: richard.booth@uni.lu
[2] Universidade da Madeira, Portugal, E-mail: ferme@uma.pt
[3] CRIL - CNRS, Lens, France, E-mail: konieczny@cril.fr
[4] Universidad de Los Andes, Mérida, Venezuela, E-mail: pino@ula.ve

[5] The proof is quite long and for space reasons we don't include it in this

the behavior of these operators. We finish in Section 6 with some concluding remarks and perspectives.

2 PRELIMINARIES

We extend the epistemic states approach of Darwiche and Pearl [9] (see also [4]). Actually we are going to give a richer notion of epistemic state. We consider, further to the beliefs, the credible formulas.

Our basic framework is finite propositional logic. This allows us to encode beliefs and credible formulas of an epistemic state by a single formula. More precisely we have the following:

Definition 1 *An* epistemic state *is an object Ψ to which we associate a consistent propositional formula $B(\Psi)$ that denotes the current beliefs of the agent in the epistemic state, and a consistent propositional formula $C(\Psi)$ that determines the credible formulas of the agent in the epistemic state and is such that $B(\Psi) \vdash C(\Psi)$.*

$C(\Psi)$ represents the credible formulas of the epistemic state Ψ. Actually, the encoding of credible formulas via $C(\Psi)$ is done as follows:

Definition 2 *α is a credible formula in the epistemic state Ψ if and only if $\alpha \wedge C(\Psi) \not\vdash \bot$.*

In previous work on (iterated) credibility-limited revision [5], standard Darwiche and Pearl epistemic states were used, and credible formulas appear as a consequence of the postulates. But we think that it is more sensible to consider the set of credible formulas as a (explicit) part of the epistemic state, since it defines how easily an agent can accept very unplausible new pieces of information. This propensity to accept unplausible pieces of information can be quite different for different agents, and defines the behavior of the agent. So it should be explicitly represented in the description of the agent, i.e., the epistemic state.

We denote the set of epistemic states by $\mathcal{E}$ and the set of consistent formulas by $\mathcal{L}^*$. We are going to consider change operators $\circ$ which are total mappings $\circ : \mathcal{E} \times \mathcal{L}^* \longrightarrow \mathcal{E}$. As usual, $\circ(\Psi, \alpha)$ is denoted by $\Psi \circ \alpha$.

In order to avoid a cumbersome treatment, we will assume the consistency of epistemic states Ψ, that is $B(\Psi) \not\vdash \bot$ and also the consistency of the new piece of information α.

Definition 3 *Given an operator $\circ$ and a natural number n, we define $\circ^n$ by recursion in the following way:*

$$\begin{aligned} \Psi \circ^0 \alpha &= \Psi \\ \Psi \circ^{n+1} \alpha &= (\Psi \circ^n \alpha) \circ \alpha \end{aligned}$$

Now we define the operator $\star$ as $\Psi \star \alpha = \Psi \circ^n \alpha$, where n is the first integer such that $B(\Psi \circ^n \alpha) \vdash \alpha$.

Note that the operator $\star$ could be partial, that is, there might exist an epistemic state Ψ and a formula α such that for every natural number n, we have $B(\Psi \circ^n \alpha) \not\vdash \alpha$. In such a case $\Psi \star \alpha$ is undefined. The fact that $\star$ is total (*i.e.* defined for all the entries in $\mathcal{E} \times \mathcal{L}^*$) will depend on properties of the operator $\circ$.

We denote by Ω the set of all interpretations. The set of models of a formula α is denoted by $[\![\alpha]\!]$. We denote by $\alpha_{\omega_1, \dots, \omega_n}$ a formula whose set of models is exactly $\{\omega_1, \dots, \omega_n\}$, i.e. s.t. $[\![\alpha_{\omega_1, \dots, \omega_n}]\!] = \{\omega_1, \dots, \omega_n\}$.

Let $\leq$ be a a total preorder, i.e., a transitive $((x \leq y \wedge y \leq z) \rightarrow x \leq z)$ and total $(x \leq y \vee y \leq x)$ relation over Ω. The corresponding strict relation $<$ is defined as $x < y$ iff $x \leq y$ and $y \not\leq x$, and the corresponding equivalence relation $\simeq$ is defined as $x \simeq y$ iff $x \leq y$ and $y \leq x$. We write $w \ll w'$ to denote when $w < w'$ and

there is no w'' such that $w < w'' < w'$. We also use the notation $min(A, \leq) = \{w \in A \mid \nexists w' \in A \; w' < w\}$.

When a set Ω is equipped with a total preorder $\leq$, then this set can be split in different levels, that gives the ordered sequence of its equivalence classes $\Omega = \langle L_0, \dots, L_n \rangle$. So $\forall x, y \in L_i$, $x \simeq y$. We say in that case that x and y are at the same level of the preorder. And $\forall x \in L_i$, $\forall y \in L_j$, if $i < j$ then $x < y$. We say in this case that x is in a lower level than y. We extend straightforwardly these definitions to compare subsets of equivalence classes, i.e if $A \subseteq L_i$ and $B \subseteq L_j$ then we say that A is in a lower level than B if $i < j$.

3 CREDIBILITY-LIMITED IMPROVEMENT

Let us first start by defining the logical properties we expect for credibility-limited improvement operators. Most of these properties are related to the ones of improvement [15, 14] and (iterated) credibility-limited revision [5]. The difficulty is to find how to obtain the wanted behavior.

We will need an additional notion (*limit of non-credibles*) for defining postulate **(CLI13)**, that describes the dynamics of credible formulas:

Definition 4 *Suppose that $\star$ is total. Given an epistemic state Ψ, such that $C(\Psi) \not\equiv \top$, we call a consistent formula λ_Ψ the* limit of non-credibles *of Ψ if the following properties hold: $\lambda_\Psi \wedge C(\Psi) \vdash \bot$ and for all β such that $\beta \wedge C(\Psi) \vdash \bot$, $B(\Psi \star (\lambda_\Psi \vee \beta)) \equiv \lambda_\Psi$.*

When there are non-credible formulas and the operator $\circ$ satisfies enough properties (see Proposition 1), this limit formula exists. This limit formula is actually the non credible formula closest to $C(\Psi)$.

Now we can give the postulates characterizing credibility-limited improvement operators. We will split them in thematic subgroups for a clearer exposition. For ease the reading we will use the following abusive shortcut: we say "revision" instead of "credibility-limited improvement", and "we revise" instead of "we perform a credibility-limited improvement". Our first, basic, group of postulates is as follows.

(CLI0) There exists an integer n such that $B(\Psi \circ^n \alpha) \vdash \alpha$
<div align="right">(Iterative success)</div>

(CLI1) If $\alpha \wedge C(\Psi) \not\vdash \bot$ then $B(\Psi \circ \alpha) \vdash \alpha$,
    else $B(\Psi \circ \alpha) \equiv B(\Psi)$ (**Relative success**)

(CLI2) If $B(\Psi) \wedge \alpha \not\vdash \bot$ then $B(\Psi \circ \alpha) \equiv B(\Psi) \wedge \alpha$
<div align="right">(Vacuity)</div>

(CLI3) $B(\Psi \circ \alpha) \not\vdash \bot$ (**Strong coherence**)

(CLI4) For all positive integers n if $\alpha_i \equiv \beta_i$ for all $i \leq n$ and $\mu \equiv \mu'$ then $B((\Psi \circ \alpha_1 \circ \cdots \circ \alpha_n) \star \mu) \equiv B((\Psi \circ \beta_1 \circ \cdots \circ \beta_n) \star \mu')$
<div align="right">(Syntax irrelevance)</div>

A first remark is that, as non-prioritized revision operators, credibility-limited revision operators do not satisfy the success postulate[6], but (CLI0) and (CLI1) are two weakenings of success. (CLI0) says that if we iterate enough we finally obtain success. This postulate has an important consequence: the corresponding $\star$ operator is total. (CLI1) uses explicitly the credible formulas[7] to decide if the new piece of information is credible enough to perform a classical AGM revision.

(CLI2), (CLI3), (CLI4) are standard revision postulates. (CLI2) is the vacuity postulate, that says that when the new piece of information is consistent with the beliefs of the agent, then the revision is just the conjunction. (CLI3) ensures that we always obtain a consistent result[8]. (CLI4) is the irrelevance of syntax condition for iteration

---

[6] (**Success**) $B(\Psi \circ \alpha) \vdash \alpha$
[7] Compared to [5] where there is no explicit set of credible formulas.
[8] Recall that we assume that the new piece of information α is always consistent.

---

paper. The interested reader can find it in [6]. However we give the main intuitions behind the proof.

(that comes from [15]).

Our second group of postulates deal with revising by conjunctions and disjunctions.

(CLI5) $B(\Psi \star \alpha) \wedge \beta \vdash B(\Psi \star (\alpha \wedge \beta))$ **(R5)**

(CLI6) If $B(\Psi \star \alpha) \wedge \beta \not\vdash \bot$, then $B(\Psi \star (\alpha \wedge \beta)) \vdash B(\Psi \star \alpha) \wedge \beta$ **(R6)**

(CLI7) $B(\Psi \circ (\alpha \vee \beta)) \equiv \begin{cases} B(\Psi \circ \alpha) \text{ or} \\ B(\Psi \circ \beta) \text{ or} \\ B(\Psi \circ \alpha) \vee B(\Psi \circ \beta) \end{cases}$ **(Trichotomy)**

It is interesting to note that (CLI5) and (CLI6) are $\star$-translations of usual properties on revision by conjunctions [15], but that we also need the trichotomy property (CLI7) on the $\circ$ operator.

The next group of postulates deal with the behavior of $B(\Psi)$ under iterated application of $\circ$.

(CLI8) If $\alpha \vdash \mu$, then $B((\Psi \circ \mu) \star \alpha) \equiv B(\Psi \star \alpha)$ **(DP1)**

(CLI9) If $\alpha \vdash \neg\mu$, then $B((\Psi \circ \mu) \star \alpha) \equiv B(\Psi \star \alpha)$ **(DP2)**

(CLI10) If $B(\Psi \star \alpha) \not\vdash \neg\mu$, then $B((\Psi \circ \mu) \star \alpha) \vdash \mu$ **(P)**

(CLI11) If $B(\Psi) \not\vdash \alpha$ then $\exists\beta$ s.t. $B((\Psi \circ \alpha) \star \beta) \not\equiv B(\Psi \star \beta)$ **(Non-triviality)**

(CLI8),(CLI9) and (CLI10) are close to standard iteration postulates. They correspond to the postulates of *rigidity* of Darwiche and Pearl called (DP1) and (DP2) [9] and to postulate (P) of Booth and Meyer [7] and Jin and Thielscher [12]. Our formulation is different because we have to define them for sequences ($\star$-version of $\circ$). (CLI8) says that starting a sequence of revisions by a less precise formula (μ) does not change the obtained beliefs. (CLI9) says that starting a sequence of revisions by a conflicting formula ($\mu \vdash \neg\alpha$) does not change the obtained beliefs. (CLI10) says that if a sequence of revision is not sufficient to imply the negation of a formula, then the increase of plausibility obtained by beginning the sequence by a revision by this formula is enough to ensure to imply it. Postulate (CLI11) says that any revision by a formula α that is not a consequence of the epistemic state modifies the epistemic state of the agent.

The next group of postulates constrain the dynamics of $C(\Psi)$ under $\circ$.

(CLI12) If $\alpha \vdash \neg\mu$ and $\alpha \wedge C(\Psi) \vdash \bot$, then $\alpha \wedge C(\Psi \circ \mu) \vdash \bot$ **(CLCD)**

(CLI13) If $\alpha \wedge C(\Psi) \vdash \bot$ and $\alpha \wedge \lambda_\Psi \not\vdash \bot$, then $\alpha \wedge C(\Psi \circ \alpha) \not\vdash \bot$ **(Enrichment of credible)**

Postulate (CLI12) says that when we revise by a formula μ then the formulas implying its negation can not enter the credible set. This postulate comes from [5]. Postulate (CLI13) says that non-credible formulas can become credible when we revise by a formula that is not credible but belongs to the limit of non-credibles.

The final postulate again deals with dynamics of $B(\Psi)$ but specifically in the non-credible revision case, so it limits the change in the epistemic state:

(CLI14) If $\mu \wedge C(\Psi) \vdash \bot$ and $B(\Psi \star \alpha) \vdash \neg\mu$ then $B((\Psi \circ \mu) \star \alpha) \not\vdash \mu$ **(Soft)**

(CLI14) imposes a limitation on the plausibility increase for non-credible formulas. It has to be compared to (CLI10). It says that if μ is not a credible formula, then the increase of plausibility caused by its revision is not enough for rejecting its negation that is obtained after some sequence of revisions (by α). This property comes from the *soft improvement* operators of [14].

Definition 5 *An operator $\circ$ satisfying CLI0-CLI14 is called a credibility-limited improvement operator.*

Let us now prove that the limit of non-credibles for credibility-limited improvement is well defined:

Proposition 1 *Suppose that the operator $\circ$ satisfies (CLI0), (CLI2-CLI6). Suppose that $C(\Psi) \not\equiv \top$. Then there exists a formula λ_Ψ satisfying the properties of Definition 4. Moreover this formula is unique up to logical equivalence.*

It is interesting to note the generality of this family of operators, since usual (admissible) iterated revision operators and (soft) improvements operators are subclasses of credibility-limited improvement operators:

Proposition 2 *Suppose that $\circ$ is a credibility-limited improvement operator. We obtain the following two special cases:*

- *If $\forall\Psi\, C(\Psi) \equiv \top$ then $\circ$ is an admissible iterated revision operator [7, 12].*
- *If $\forall\Psi\, C(\Psi) \equiv B(\Psi)$ then $\circ$ is a soft improvement operator [14].*

4 REPRESENTATION THEOREM

Let us now give a representation theorem for credibility-limited improvement operators in terms of plausibility preorders on interpretations (faithful assignments [13, 9]).

An assignment is a function mapping epistemic states into total preorders over Ω. The assignments are denoted $\Psi \mapsto \leq_\Psi$, which means, as usual, that the image of the epistemic state Ψ under the assignment is the total preorder $\leq_\Psi$.

Given an assignment and a change operator $\circ$ we define a number of properties. First, regarding the relationship between $\leq_\Psi$ and $B(\Psi)$:

(SCLI1) If $\omega \models B(\Psi)$ and $\omega' \models B(\Psi)$, then $\omega \simeq_\Psi \omega'$

(SCLI2) If $\omega \models B(\Psi)$ and $\omega' \not\models B(\Psi)$, then $\omega <_\Psi \omega'$

Conditions (SCLI1) and (SCLI2) just say that the models of $B(\Psi)$ are the minimal elements of $\leq_\Psi$. The next postulates are about the relationship between $\leq_\Psi$ and $\leq_{\Psi \circ \alpha}$

(SCLI3) For all positive integers n if $\alpha_i \equiv \beta_i$ for all $i \leq n$ then $\leq_{\Psi \circ \alpha_1 \circ \cdots \circ \alpha_n} = \leq_{\Psi \circ \beta_1 \circ \cdots \circ \beta_n}$

(SCLI4) If $\omega, \omega' \models \alpha$ then $\omega \leq_\Psi \omega' \Leftrightarrow \omega \leq_{\Psi \circ \alpha} \omega'$

(SCLI5) If $\omega, \omega' \models \neg\alpha$ then $\omega \leq_\Psi \omega' \Leftrightarrow \omega \leq_{\Psi \circ \alpha} \omega'$

(SCLI6) If $\omega \models \alpha$, $\omega' \models \neg\alpha$ and $\omega \leq_\Psi \omega'$, then $\omega <_{\Psi \circ \alpha} \omega'$

(SCLI7) If $min(\leq_\Psi) \not\subseteq [\![\alpha]\!]$ then $\leq_\Psi \neq \leq_{\Psi \circ \alpha}$

(SCLI3) is the semantical counterpart of irrelevance of the syntax. (SCLI4) and (SCLI5) are the conditions of rigidity: the relative order of models of α is preserved after revision; the same happens with the models of $\neg\alpha$. (SCLI6) guarantees that the plausibility of models of the new information is improved with respect to the countermodels of this information after revision (see Proposition 3). So if a model and a countermodel of α have the same plausibility, after revision the model will be more plausible than the countermodel. (SCLI7) guarantees that something changes after revision if the current beliefs doesn't already contain the new piece of information.

Our last group of postulates deal with how $C(\Psi)$ and its dynamics under $\circ$ are reflected in the assignment $\Psi \mapsto \leq_\Psi$.

(SCLI8) If $\omega \models C(\Psi)$ and $\omega' \not\models C(\Psi)$ then $\omega <_\Psi \omega'$

(SCLI9) If $\omega \ll_\Psi \omega'$, $\omega \models C(\Psi)$, $\omega' \in [\![\alpha]\!]$ and $[\![\alpha]\!] \cap [\![C(\Psi)]\!] = \emptyset$ then $\omega' \models C(\Psi \circ \alpha)$

(SCLI10) If $\omega \models \neg\alpha$ and $\omega \not\models C(\Psi)$ then $\omega \not\models C(\Psi \circ \alpha)$

(SCLI11) If $[\![\alpha]\!] \cap [\![C(\Psi)]\!] = \emptyset$, $\omega \models \alpha, \omega' \models \neg\alpha$ then $\omega' <_\Psi \omega \Rightarrow \omega' \leq_{\Psi \circ \alpha} \omega$

(SCLI8) says that credible interpretations are more plausible than non-credible interpretations. (SCLI9) says that a non-credible model of revision formula α becomes credible after revision if it is amongst the most plausible non-credible interpretations. (SCLI10) forbids the models of the negation of the revision formula to become credible after the revision. (SCLI11) expresses that the plausibility change is a small (soft) one for the interpretations of a non-credible new piece of information.

Definition 6 *An assignment $\Psi \mapsto \leq_\Psi$ satisfying the properties SCLI1-SCLI11 for the operator $\circ$ is called a CLI-faithful assignment for $\circ$.*

Note that the condition (SCLI6) (which was proposed in [7, 12]) implies both of the conditions (CR3) and (CR4) of Darwiche and Pearl [9]:

(CR3) If $\omega \models \alpha, \omega' \models \neg\alpha$ then $\omega <_\Psi \omega' \Rightarrow \omega <_{\Psi \circ \alpha} \omega'$
(CR4) If $\omega \models \alpha, \omega' \models \neg\alpha$ then $\omega \leq_\Psi \omega' \Rightarrow \omega \leq_{\Psi \circ \alpha} \omega'$

These conditions are important in order to guarantee that there is no loop when we iterate the revisions by α and that after enough iterations of α the minimal elements of the associated preorder are models of α. Let us give a proposition that will be useful to show that property:

Proposition 3 *If $\Psi \mapsto \leq_\Psi$ is a CLI-faithful assignment for $\circ$, then the following condition holds:*

(SM) *If $\alpha \nvdash \bot$ and $[\![\neg\alpha]\!] \cap min(\leq_\Psi) \neq \emptyset$ then at least one of the following conditions holds:*

   (i) $\exists \omega_1, \omega_2$ *s.t.* $\omega_1 \models \alpha, \omega_2 \models \neg\alpha, \omega_1 \simeq_\Psi \omega_2$ *and* $\omega_1 <_{\Psi \circ \alpha} \omega_2$

   (ii) $\exists \omega_1, \omega_2$ *s.t.* $\omega_1 \models \alpha, \omega_2 \models \neg\alpha, \omega_2 <_\Psi \omega_1$ *and* $\omega_1 \leq_{\Psi \circ \alpha} \omega_2$

This proposition expresses the fact that when the beliefs do not imply a formula α, the plausibility of at least one model of α is improved after revision: either a model that was equivalent to a countermodel is now strictly more plausible, or a model that was strictly less plausible than a countermodel is now equivalent to this countermodel.

Let us now give the representation theorem for credibility-limited improvement operators.

Theorem 1 *Let $\circ$ be a change operator. The operator $\circ$ is a credibility-limited improvement operator only if there is a CLI-faithful assignment for $\circ$, $\Psi \mapsto \leq_\Psi$, such that the following conditions hold:*

(i) *If $\alpha \wedge C(\Psi) \nvdash \bot$ then $[\![B(\Psi \circ \alpha)]\!] = min([\![\alpha]\!], \leq_\Psi)$.*
(ii) *If $\alpha \wedge C(\Psi) \vdash \bot$ then $[\![B(\Psi \circ \alpha)]\!] = [\![B(\Psi)]\!]$.*
(iii) *For all α, $[\![B(\Psi \star \alpha)]\!] = min([\![\alpha]\!], \leq_\Psi)$.*

Conversely, suppose we have a CLI-faithful assignment for $\circ$ such that the conditions (i) and (ii) above are satisfied. Then the operator $\circ$ is a credibility-limited improvement operator and condition (iii) is satisfied.

The detailed proof of this result is quite long and for space reasons cannot be included here. The interested reader can find it in [6]. However, we give below the main lines of reasoning.

Sketch of the proof of Theorem 1: *From the syntactical side to the semantical side.* Assume that $\circ$ is a credibility-limited improvement operator. For each epistemic state Ψ define a binary relation $\leq_\Psi$ in the following way: $\omega \leq_\Psi \omega' \Leftrightarrow \omega \models B(\Psi \star \alpha_{\omega,\omega'})$. First we prove that $\leq_\Psi$ is a total preorder. This part of the proof, using the postulates (CLI0) and (CLI2-CLI6), follows the techniques in [15]. In order to prove that the assignment $\Psi \mapsto \leq_\Psi$ is indeed a CLI-faithful assignment for $\circ$, we verify each condition of Definition 6. Conditions (SCLI1) and (SCLI2) follow from (CLI2). Condition (SCLI3) follows from (CLI4). With postulates (CLI0) and (CLI2-CLI6) and their derivable condition (SCLI1) we prove condition (iii) of the Theorem, *i.e.* $[\![B(\Psi \star \alpha)]\!] = min([\![\alpha]\!], \leq_\Psi)$.

We continue the sketch of verification of the other conditions of CLI-faithful assignment. Condition (SCLI4) follows from (CLI8) and condition (iii). Condition (SCLI5) follows from (CLI9) and condition (iii). Condition (SCLI6) follows from (CLI0), (CLI3), (CLI10) and condition (iii). Condition (SCLI7) follows from (CLI11) and Condition (iii). Condition (SCLI8) follows from (CLI1), (CLI7) and condition (iii). Condition (SCLI9) follows from (CLI13), (SCLI4), (SCLI8) and condition (iii). Condition (SCLI10) follows from (CLI12). Condition (SCLI11) follows from (CLI14) and condition (iii).

It remains to prove condition (i) and condition (ii). First we verify that condition (i) holds. Suppose $\alpha \wedge C(\Psi) \nvdash \bot$. Then, by (CLI1), $B(\Psi \circ \alpha) \vdash \alpha$. Thus, by definition of $\star$, $B(\Psi \circ \alpha) = B(\Psi \star \alpha)$. Then, by condition (iii), $[\![B(\Psi \circ \alpha)]\!] = min([\![\alpha]\!], \leq_\Psi)$, that is, condition (i) holds. Condition (ii) follows straightforwardly from (CLI1).

From the semantical side to the syntactical side. The first step will be to prove postulate (CLI0) and then that property (iii) holds. In order to do this first task, we adapt some techniques introduced recently in [16]. This is the most elaborate part of the proof. After that and once property (iii) is established, checking that the other postulates hold is a relatively painless task. The main idea in order to prove (CLI0) is very intuitive. Note that to prove (CLI0) is equivalent to prove that there exists a positive integer i such that $[\![B(\Psi \circ^i \alpha)]\!] \subseteq [\![\alpha]\!]$. We associate to each model ω of α the number $\ell^i(\omega)$ which is the number of levels in the preorder $\leq_{\Psi \circ^i \alpha}$ such that there exists $\omega' \in [\![\neg\alpha]\!]$ satisfying $\omega' \leq_{\Psi \circ^i \alpha} \omega$. Then we associate to α the vector $v_i(\alpha)$ consisting of the numbers $\ell^i(\omega)$ for each $\omega \models \alpha$ ordered increasingly. It is easy to see that $[\![B(\Psi \circ^i \alpha)]\!] \subseteq [\![\alpha]\!]$ if, and only if, the first coordinate of vector $v_i(\alpha)$ is 0 (actually in such a case $[\![B(\Psi \circ^i \alpha)]\!]$ is the set $\{\omega \in [\![\alpha]\!] : \ell^i(\omega) = 0\}$). What the conditions of CLI-faithful assignment allow to prove is that, as long as the first coordinate of the vector $v_i(\alpha)$ is different from 0, then there exists an integer $j > 0$ such that $v_i(\alpha) >_{lex} v_{i+j}(\alpha)$, where $>_{lex}$ is the strict lexicographic order. Finally, since the strict lexicographic order is well founded, necessarily there exists an integer n such that the first coordinate of the vector $v_n(\alpha)$ is equal to 0. From the conditions of *rigidity* it is not hard to see that for all i, $min([\![\alpha]\!], \leq_\Psi) = min([\![\alpha]\!], \leq_{\Psi \circ^i \alpha})$ and from this and the previous discussion it is relatively easy to see that condition (iii) holds.

It remains to check postulates (CLI1-CLI14). Here, we give the conditions involved in the proof of each postulate. Postulate (CLI1) follows from conditions (i) and (ii). Postulate (CLI2) follows from conditions (SCLI1) and (SCLI2). Postulate (CLI3) is trivial by our assumptions of consistency of both α and $B(\Psi)$. Postulate (CLI4) follows from conditions (SCLI3) and (iii). Postulates (CLI5) and (CLI6) follow from condition (iii). Postulate (CLI7) follows from conditions (i) and (SCLI8). Postulate (CLI8) follows from conditions (SCLI4) and (iii). Postulate (CLI9) follows from conditions (SCLI5) and (iii). Postulate (CLI10) follows from conditions (SCLI6) and (iii). Postulate (CLI12) follows from condition (SCLI10). Postulate (CLI11) follows from conditions (SCLI7) and (iii). Postulate

(CLI13) follows from condition (SCLI9). Postulate (CLI14) follows from conditions (SCLI11) and (iii). ∎

We have to note that this representation theorem has some similarities and differences with the usual ones for iterated revision [9, 7, 15, 5]. The part leading from syntactical postulates to the assignment is quite similar to the same part in the representation theorems of previous works. However the converse, unlike those previous representation theorems, requires a more elaborate and complex proof. This is due to the fact that we want to merge the behavior of (admissible) iterated revision operators and of (soft) improvement operators. In particular, we have to face a double difficulty. First, we don't have the ⋆ operator because we don't know, a priori, if the postulate of Iterated success (CLI0) holds. Second, and as consequence of the first difficulty, we can't assume condition (iii). The solution of this double difficulty is what was mainly sketched above.

5 ILLUSTRATIVE EXAMPLE

Let us define a concrete credibility-limited improvement operator in order to illustrate their behavior.

The idea is to use a particular improvement operator, the *one-improvement* operator [14] in order to increase the plausibility of the new piece of information. We will use this operator for both credible and non-credible formulas. But for credible formulas we have to make a revision, so after the one-improvement we will apply Boutilier's *natural revision* operator [8], which will ensure that the most plausible models of the new piece of information become the most plausible models for the agent. Let us use • to denote this operator, which is defined intuitively as follows.

Definition 7 *Let $\circ_N$ be Boutilier's natural revision operator [8], and let $\odot$ be the one-improvement operator [14]*

$$\Psi \bullet \alpha = \begin{cases} \Psi \odot \alpha & \text{If } \alpha \wedge C(\Psi) \vdash \bot \\ (\Psi \odot \alpha) \circ_N \alpha & \text{Otherwise} \end{cases}$$

In order to make formal the previous definition in our setting, we have to define the credible formulas of each epistemic state, which indeed are absent for natural revision as well as for one-improvement operators.

Let us see now how to define this operator directly using the semantical definition of the credibility-limited operators. We will use a concrete representation of the epistemic states: total preorders over interpretations. So suppose from here that an epistemic state Ψ is represented by a preorder $\leq_\Psi$ and a set of credible interpretations C_Ψ formed by the the union of the first k_Ψ levels $L_0, \dots, L_{k_\Psi}$ of the preorder $\leq_\Psi$, that is $C_\Psi = \cup_{j=0}^{k_\Psi} L_j$ and $\Psi = (\leq_\Psi, C_\Psi)$. This is an acceptable representation of epistemic state. In this setting, $B(\Psi)$ and $C(\Psi)$ are formulas whose set of models are respectively $\min(\leq_\Psi)$ and C_Ψ.

So now we define the • CLI operator as a transition function between these epistemic states. Take any epistemic state $\Psi = (\leq_\Psi, C_\Psi)$, and any formula α, and let us define $\Psi \bullet \alpha = (\leq_{\Psi \bullet \alpha}, C_{\Psi \bullet \alpha})$. We must then specify $\leq_{\Psi \bullet \alpha}$ and $C_{\Psi \bullet \alpha}$.

Turning first to $\leq_{\Psi \bullet \alpha}$, as indicated above this will be defined in terms of one-improvement and natural revision. The definition of the preorder $\leq_{\Psi \odot \alpha}$ given by one-improvement is as follows.

Definition 8 *Let $\Psi = (\leq_\Psi, C_\Psi)$ be an epistemic state and α be a formula.*

- *If $\omega, \omega' \in \llbracket \alpha \rrbracket$ then $\omega \leq_\Psi \omega' \Leftrightarrow \omega \leq_{\Psi \odot \alpha} \omega'$*
- *If $\omega, \omega' \in \llbracket \neg \alpha \rrbracket$ then $\omega \leq_\Psi \omega' \Leftrightarrow \omega \leq_{\Psi \odot \alpha} \omega'$*
- *If $\omega \in \llbracket \alpha \rrbracket, \omega' \in \llbracket \neg \alpha \rrbracket$ then*

$\qquad$ – $\omega <_\Psi \omega' \Rightarrow \omega <_{\Psi \odot \alpha} \omega'$
$\qquad$ – $\omega \simeq_\Psi \omega' \Rightarrow \omega <_{\Psi \odot \alpha} \omega'$
$\qquad$ – $\omega' \ll_\Psi \omega \Rightarrow \omega' \simeq_{\Psi \odot \alpha} \omega$
$\qquad$ – $\omega' <_\Psi \omega$ *and* $\omega' \not\ll_\Psi \omega \Rightarrow \omega' <_{\Psi \odot \alpha} \omega$

The next proposition ensures that the preorder transition function in the above definition is well-defined:

Proposition 4 *There is a single relation $\leq_{\Psi \odot \alpha}$ that satisfies the conditions in Definition 8. Moreover $\leq_{\Psi \odot \alpha}$ is a total preorder.*

Now let us define Boutilier's natural revision on this representation. The idea is to keep the same preorder, except that the most plausible models of the formula α become the most plausible models of the epistemic state:

Definition 9 *Let $\Psi = (\leq_\Psi, C_\Psi)$ be an epistemic state and α be a formula, then the definition of the preorder $\leq_{\Psi \circ_N \alpha}$ given by the natural revision operator is as follows.:*

- *If $\omega, \omega' \in min(\llbracket \alpha \rrbracket, \leq_\Psi)$ then $\omega \leq_{\Psi \circ_N \alpha} \omega' \Leftrightarrow \omega \leq_\Psi \omega'$*
- *If $\omega, \omega' \notin min(\llbracket \alpha \rrbracket, \leq_\Psi)$ then $\omega \leq_{\Psi \circ_N \alpha} \omega' \Leftrightarrow \omega \leq_\Psi \omega'$*
- *If $\omega \in min(\llbracket \alpha \rrbracket, \leq_\Psi), \omega' \notin min(\llbracket \alpha \rrbracket, \leq_\Psi)$ then $\omega <_{\Psi \circ_N \alpha} \omega'$*

Now we can define $\leq_{\Psi \bullet \alpha}$:

Definition 10 *Let $\Psi = (\leq_\Psi, C_\Psi)$ be an epistemic state and α be a formula.*

$$\leq_{\Psi \bullet \alpha} = \begin{cases} \leq_{\Psi \odot \alpha} & \text{if} \quad \llbracket \alpha \rrbracket \cap C_\Psi = \emptyset \\ \leq_{(\Psi \odot \alpha) \circ_N \alpha} & \text{otherwise} \end{cases}$$

Let us turn now to the evolution of credible interpretations:

Definition 11 *Let $\Psi = (\leq_\Psi, C_\Psi)$ be an epistemic state and α be a formula. Define $C_{\Psi \bullet \alpha}$ as $C_{\Psi \bullet \alpha} = C_\Psi \cup \{\omega \in \llbracket \alpha \rrbracket : \omega \notin C_\Psi$ and $\exists \omega' \in C_\Psi$ $\omega' \ll_\Psi \omega\}$.*

We are now able to define the • CLIO operator in our concrete framework:

Definition 12 *Let $\Psi = (\leq_\Psi, C_\Psi)$ be an epistemic state and α be a formula. $\Psi \bullet \alpha = (\leq_{\Psi \bullet \alpha}, C_{\Psi \bullet \alpha})$*

As expected, with this concrete representation of epistemic states, the definition of the • operator adheres to the intuitive definition given in Definition 7. And of course this operator is built as an example of a credibility-limited improvement operator:

Proposition 5 *The operator • is a credibility-limited improvement operator, i.e. it satisfies (CLI0-CLI14).*

This operator can also be defined using a representation close to Ordinal Conditional Functions (see [15]).

Let us now illustrate the behavior of this credibility-limited revision operator on a concrete example.

Example 2 *Bob is a citizen of a country under a dictatorship who is skeptical about the efficacy of demonstrations in order to obtain a change towards more freedom and democracy. After the positive changes in Cosivia and Guyostan which followed some demonstrations, he is less skeptical and thinks that it might be plausible that demonstrations produce changes towards more freedom and democracy. After further positive events in Austropia, he is finally convinced that demonstrations can make changes to more freedom or more democracy or both. This example can be modelled by the following logical representation: we take three propositional variables, p, f and d in this order, encoding respectively* demonstrations *(protests),* more freedom *and* more democracy. *The formula α representing the new information is $p \wedge (p \to f \vee d)$. Thus, $\llbracket \alpha \rrbracket = \{101, 110, 111\}$. The initial epistemic state $\Psi = (\leq_\Psi, C_\Psi)$ is represented as follows (in black the models of C_Ψ and in gray the other interpretations (that are not credible), the models of α are encircled):*
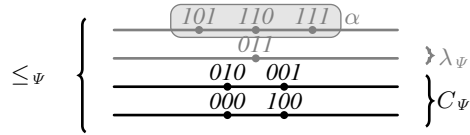
The credible worlds C_Ψ are the first two levels, i.e. $C_\Psi = \{010, 001, 000, 100\}$, and $[\![B(\Psi)]\!] = \{000, 100\}$. The sequence of iterations $\Psi \bullet \alpha$, $\Psi \bullet^2 \alpha$ and $\Psi \bullet^3 \alpha$ is represented as follows:
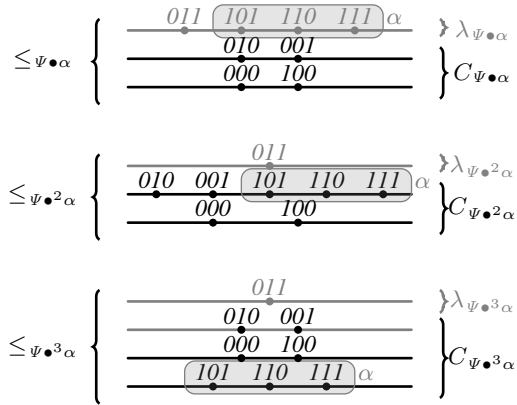
Note that $C_\Psi = C_{\Psi \bullet \alpha}$, then the models of α enter the set of credible interpretations during $\bullet^2$. Then $C_{\Psi \bullet^2 \alpha} = C_{\Psi \bullet^3 \alpha}$ but $C_{\Psi \bullet^2 \alpha}$ is formed by the first two levels of $\leq_{\Psi \bullet^2 \alpha}$ whereas $C_{\Psi \bullet^3 \alpha}$ is formed by the first three levels of $\leq_{\Psi \bullet^3 \alpha}$. The epistemic state $\Psi \bullet \alpha$ represents the state of the citizen after the events in Cosivia. The epistemic state $\Psi \bullet^2 \alpha$ represents the state of the citizen after the events in (Cosivia and) Guyostan. Note that at this step α becomes credible, which means that at the next iteration α will be accepted by the agent (i.e. the revision will be successful). So finally, the epistemic state $\Psi \bullet^3 \alpha$ represents the state of the citizen after the events in (Cosivia, Guyostan and) Austropia, where the agent makes a revision by this credible formula, and finally believes it.

6 CONCLUSION AND DISCUSSION

In this paper, we proposed a new model of belief change, credibility-limited improvement operators, based on credibility-limited revision and improvement operators. The outcome is a large family of operators, which have the particularity of being non-prioritized and well-behaved with respect to iteration. The new information is either accepted or gains in plausibility in the epistemic state of the agent.

As usual for iterated revision operators, several different representation frameworks can be used to encode credibility-limited improvement operators, such as Ordinal Conditional Functions [17], or conditionalizations in possibilistic logic bases [3, 2]. Even if none of the existing operators on these representations can directly encode credibility-limited improvement operators, some combinations of operators can achieve this. However there is no logical characterization of operators in these settings.

Our distinction between non-credible and credible interpretations can be related to the distinction between impossible and plausible interpretations in possibility distributions. The difference is that in possibility distributions all the impossible interpretations are equivalent, whereas we still rank the non-credible interpretations, that can be sensible (as advocated in [**?**]) if we consider the dichotomy as reflecting a (strong) difference of certainty, and not as a knowledge/belief frontier like in possibilistic logic. So possibility distributions can be seen as a particular case of our epistemic states.

REFERENCES

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson, 'On the logic of theory change: Partial meet contraction and revision functions', *Journal of Symbolic Logic*, **50**, 510–530, (1985).

[2] Salem Benferhat, Didier Dubois, Henri Prade, and Mary-Anne Williams, 'A practical approach to revising prioritized knowledge bases', *Studia Logica*, **70**(1), 105–130, (2002).

[3] Salem Benferhat, Didier Dubois, Henri Prade, and Mary-Anne Williams, 'A framework for iterated belief revision using possibilistic counterparts to Jeffrey's rule', *Fundamenta Informaticae*, **99**(2), 147–168, (2010).

[4] Salem Benferhat, Sébastien Konieczny, Odile Papini, and Ramon Pino Pérez, 'Iterated revision by epistemic states: axioms, semantics and syntax', in *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI'2000)*, pp. 13–17, (2000).

[5] Richard Booth, Eduardo Fermé, Sébastien Konieczny, and Ramón Pino Pérez, 'Credibility-limited revision operators in propositional logic', in *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012)*, pp. 116–125, (2012).

[6] Richard Booth, Eduardo Fermé, Sébastien Konieczny, and Ramón Pino Pérez, 'Credibility-limited improvement operators', URL: http://www.cril.fr/~konieczny/clio-proofs.pdf, (2014).

[7] Richard Booth and Thomas Meyer, 'Admissible and restrained revision', *Journal of Artificial Intelligence Research*, **26**, 127–151, (2006).

[8] Craig Boutilier, 'Revision sequences and nested conditionals', in *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, pp. 519–525, (1993).

[9] Adnan Darwiche and Judea Pearl, 'On the logic of iterated belief revision', *Artificial Intelligence*, **89**, 1–29, (1997).

[10] Eduardo Fermé and Sven Ove Hansson, 'AGM 25 years: Twenty-five years of research in belief change', *Journal of Philosophical Logic*, **40**, 295–331, (2011).

[11] Sven Ove Hansson, Eduardo Fermé, John Cantwell, and Marcelo Falappa, 'Credibility limited revision', *Journal of Symbolic Logic*, **66**, 1581–1596, (2001).

[12] Yi Jin and Michael Thielscher, 'Iterated belief revision, revised', *Artificial Intelligence*, **171**, 1–18, (2007).

[13] Hirofumi Katsuno and Alberto O. Mendelzon, 'Propositional knowledge base revision and minimal change', *Artificial Intelligence*, **52**, 263–294, (1991).

[14] Sébastien Konieczny, Mattia Medina Grespan, and Ramón Pino Pérez, 'Taxonomy of improvement operators and the problem of minimal change', in *Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, pp. 161–170, (2010).

[15] Sébastien Konieczny and Ramón Pino Pérez, 'Improvement operators', in *Principles of the 11th International Conference on Principles of Knowledge Representation and Reasoning (KR 2008)*, pp. 177–186, (2008).

[16] Mattia Medina Grespan and Ramón Pino Pérez, 'Representation of basic improvement operators', in *Trends in Belief Revision and Argumentation Dynamics*, 195–227, College Publications, (2013).

[17] Wolfgang Spohn, 'Ordinal conditional functions: A dynamic theory of epistemic states', in *Causation in Decision, Belief Change, and Statistics*, volume 2, 105–134, (1987).