Explanation for KLM-Style Defeasible Reasoning

Lloyd Everett[1]0000-0002-8706-831X, Emily Morris[1]0000-0003-1336-3903, and Thomas Meyer[1,2]0000000322046969

[1] University of Cape Town, Cape Town, South Africa
[2] Centre for Artificial Intelligence Research, South Africa

Abstract. Explanation services are a crucial aspect of symbolic reasoning systems but they have not been explored in detail for defeasible formalisms such as KLM. We evaluate prior work on the topic with a focus on KLM propositional logic and find that a form of defeasible explanation initially described for Rational Closure which we term *weak justification* can be adapted to Relevant and Lexicographic Closure as well as described in terms of intuitive properties derived from the KLM postulates. We also consider how a more general definition of defeasible explanation known as strong explanation applies to KLM and propose an algorithm that enumerates these justifications for Rational Closure.

Keywords: knowledge representation and reasoning · defeasible reasoning · KLM approach · Rational Closure · Relevant Closure · Lexicographic Closure · explanations

1 Introduction

Explanation services indicate to users of symbolic reasoning systems which parts of their knowledge base lead to particular conclusions. This is helpful particularly when the reasoner is giving unexpected results since it allows the user to identify the culprit knowledge base statements and thus debug their knowledge base [10]. Explanation services have also been found to improve knowledge base comprehension, particularly if the user is not familiar with the knowledge base [1], and to improve users' confidence in the reasoning system [2]. There is also some evidence that formalisms of explanation can be theoretical tools in their own right; for example, Casini et al. [4] base their work on Relevant Closure fundamentally on classical justification, a form of classical explanation.

Although well-understood in the classical case, explanation has not yet been explored in detail for defeasible reasoning apart from some foundational work [8, 3]. Our work aims to improve our understanding of explanation for defeasible propositional logic and where relevant to provide algorithms for the practical implementation of explanation services.

There are many approaches to defeasible reasoning but a particularly compelling approach that has been studied at length in the literature [5, 7, 14, 13, 6] is the KLM approach suggested by Kraus, Lehmann and Magidor [11]. One of the major appeals of KLM is that it can be viewed from two different angles, each with its own advantages: either using a series of postulates asserting behaviours we intuitively expect of the defeasible reasoning formalism, or using a model-theoretic semantics perhaps not as obviously intuitive but more amenable to computation by means of reasoning algorithms. These two perspectives are linked by results in the literature [13, 9, 6]. Formalisms of defeasible entailment explored in the literatue for KLM include *Rational Closure* [13], *Relevant Closure* [4] and *Lexicographic Closure* [12].

Chama [8] proposes an algorithm for the evaluation of defeasible justifications for Rational Closure. We term this notion of defeasible justification *weak justification* and adapt this result to the cases of Relevant Closure and Lexicographic Closure. We then consider how this notion relates to *strong explanation*, a more general notion of defeasible explanation given by Brewka and Ulbricht [3] which has not yet been explored for KLM, and propose an algorithm for enumerating strong justifications for the case of Rational Closure using a revised definition of strong justification. Our final result characterises weak justification using properties with intuitive interpretations based on the KLM postulates.

2 Background

2.1 Classical Propositional Logic

We begin with a finite set $\mathcal{P} = \{p, q, \cdots\}$ of *propositional atoms*. The binary connectives $\wedge, \vee, \rightarrow, \leftrightarrow$ and the negation operator $\neg$ are defined recursively to form propositional *formulas*. The set of all such formulas over $\mathcal{P}$ is the *propositional language* $\mathcal{L}$. A *valuation* is a function $\mathcal{P} \rightarrow \{\mathrm{T}, \mathrm{F}\}$ that assigns a truth value to each atom in $\mathcal{P}$. We say that a formula $\alpha \in \mathcal{L}$ is *satisfied* by a valuation $\mathcal{I}$ if α evaluates to true according to the usual truth-functional semantics given $\mathcal{I}$. The valuations that satisfy a formula α are referred to as *models* of α, and the set of models of α is denoted $\mathrm{Mod}\,(\alpha)$. By assertion, $\top$ is a propositional formula satisfied by every valuation and $\bot$ is a formula not satisfied by any valuation.

A *classical knowledge base* $\mathcal{K}$ is a finite set of propositional formulas. A valuation is a model of $\mathcal{K}$ if it is a model of every formula in $\mathcal{K}$. A knowledge base $\mathcal{K}$ *entails* a formula α, denoted $\mathcal{K} \models \alpha$, if $\mathrm{Mod}\,(\mathcal{K}) \subseteq \mathrm{Mod}\,(\alpha)$ and a formula α entails a formula β, denoted $\alpha \models \beta$, if $\mathrm{Mod}\,(\alpha) \subseteq \mathrm{Mod}\,(\beta)$. A knowledge base $\mathcal{J}$ is a *justification* for an entailment $\mathcal{K} \models \alpha$ if $\mathcal{J}$ is a subset $\mathcal{J} \subseteq \mathcal{K}$ such that $\mathcal{J} \models \alpha$ and there is no proper subset $\mathcal{J}' \subset \mathcal{J}$ such that $\mathcal{J}' \models \alpha$. Algorithms for enumerating classical justifications have been explored in detail by Horridge [10].

2.2 KLM Defeasible Entailment

Although there are many approaches to defeasible reasoning, one approach that has been studied extensively in the literature is that proposed by Kraus, Lehmann and Magidor (KLM) [11]. This approach extends propositional logic by introducing defeasible implication $\mid\!\sim$ which can be viewed as the defeasible analogue of classical implication $\rightarrow$. Defeasible implications (DI) are expressions of the form $\alpha \mid\sim \beta$ where $\alpha \in \mathcal{L}, \beta \in \mathcal{L}$ and are read as 'α *typically implies* β'.

A defeasible knowledge base is then a finite set of defeasible implications and defeasible entailment $\approx$ is defined as a binary relation over defeasible knowledge baes and defeasible implications so that $\mathcal{K} \approx \alpha \mid\sim \beta$ reads as '$\mathcal{K}$ *defeasibly entails that* α *typically implies* β'. Note that while we assume that defeasible knowledge bases only contain defeasible implications, we can express any classical formula α using the defeasible representation $\neg\alpha \mid\sim \bot$. From here on we will assume that knowledge bases are defeasible unless stated otherwise.

Lehmann and Magidor [13] propose a series of postulates that define *rational* defeasible entailment, where each postulate can be thought of as asserting an intuitive characteristic we expect of a sensible defeasible entailment relation (hence the name *rational*). In addition to this axiomatic definition, rational entailment relations have a model-theoretic semantics which we do not discuss here but which is (in some cases) described exactly by reasoning algorithms of reasonable computational complexity. These reasoning algorithms are a central focus in this paper and we introduce Rational Closure [13], the most well-known form of defeasible entailment for KLM, in these terms.

2.3 Rational Closure

Rational Closure is a rational definition for defeasible entailment proposed by Lehmann and Magidor [13]. Casini et al. [6] present an algorithm for Rational Closure with two distinct sub-phases, shown in Algorithms 1 and 2. Essentially, the algorithm works by imposing a ranking of typicality on the knowledge base. Then, if there is an inconsistency when computing entailment, the most typical information in this ranking is removed from the knowledge base. The ranking of statements is produced by `BaseRank`, shown in Algorithm 2. The lower the rank of a statement, the more typical it is.

Algorithm 1: RationalClosure

Input: A knowledge base $\mathcal{K}$ and a DI $\alpha \mid\sim \beta$
Output: **true**, if $\mathcal{K} \approx_{RC} \alpha \mid\sim \beta$, otherwise **false**

$(R_0, R_1, ..., R_\infty$, n) := BaseRank(K);
i := 0;
$R := \bigcup_{i=0}^{j<n} R_j$;
while $R_\infty \cup R \models \neg\alpha$ ***and*** $R \neq \emptyset$ **do**
  $R := R \setminus R_i$;
  i := i+1;
end
return $R_\infty \cup R \models \alpha \rightarrow \beta$;

As an illustration of the Rational Closure algorithm, consider the following example.

Example 1. Suppose one has the defeasible knowledge base $\mathcal{K}$ containing the following information.

Algorithm 2: BaseRank

Input: A knowledge base $\mathcal{K}$

Output: An ordered tuple $(R_0, ..., R_{n-1}, R_\infty$, n)

i := 0;

$E_0 := \overline{\mathcal{K}}$;

repeat

  $E_{i+1} := \{\alpha \rightarrow \beta \in E_i \mid E_i \models \neg\alpha\}$;

  $R_i := E_i \setminus E_{i+1}$;

  i := i+1;

until $E_{i-1} = E_i$;

$R_\infty := E_{i-1}$;

n := i-1;

return *$(R_0, ..., R_{n-1}, R_\infty$, n)*;

1. Birds typically fly ($b \mid\!\sim f$)
2. Birds typically have eyes ($b \mid\!\sim e$)
3. Birds typically sing ($b \mid\!\sim s$)
4. Penguins typically do not fly ($p \mid\!\sim \neg f$)
5. Penguins are birds ($p \rightarrow b$)
6. Max is a penguin ($m \rightarrow p$)

Consider the entailment of the statement 'Max typically does not fly' ($m \mid\!\sim \neg f$). Using `RationalClosure`, `BaseRank` is first used to compute the ranking in Figure 1. Then we start by considering all the ranks and check whether $R_0 \cup R_1 \cup R_\infty \models \neg m$. Since this holds, R_0 is removed. We then check whether $R_1 \cup R_\infty \models \neg m$. Since this entailment does not hold, we stop removing ranks and check whether $R_1 \cup R_\infty \models m \rightarrow \neg f$. Since this entailment holds, `RationalClosure` will return **true**.

0	$b \mid\!\sim f, b \mid\!\sim e, b \mid\!\sim s$
1	$p \mid\!\sim \neg f$
∞	$p \rightarrow b, m \rightarrow p$

Fig. 1: Base Ranking of statements for Example 1

It is helpful to introduce some notation closely related to these algorithms [13]:

Definition 1. *The materialisation $\overline{\mathcal{K}}$ of a knowledge base $\mathcal{K}$ is the classical knowledge base $\{\alpha \rightarrow \beta \mid \alpha \mid\!\sim \beta \in \mathcal{K}\}$.*

Definition 2. *The exceptionality sequence $\mathcal{E}_0^{\mathcal{K}}, \ldots, \mathcal{E}_n^{\mathcal{K}}$ for a knowledge base $\mathcal{K}$ is given by letting $\mathcal{E}_0^{\mathcal{K}} = \mathcal{K}$, and $\mathcal{E}_{i+1}^{\mathcal{K}} = \left\{\alpha \mid\!\sim \beta \in \mathcal{E}_i^{\mathcal{K}} \mid \overline{\mathcal{E}_i^{\mathcal{K}}} \models \neg\alpha\right\}$ for $0 \leq i < n$ where n is the smallest index such that $\mathcal{E}_n^{\mathcal{K}} = \mathcal{E}_{n+1}^{\mathcal{K}}$ according to these equations. The final element $\mathcal{E}_n^{\mathcal{K}}$ is usually denoted as $\mathcal{E}_\infty^{\mathcal{K}}$ as it is unique in that its statements are never retracted when evaluating entailment queries.*

Definition 3. *The base rank* $\mathrm{br}_{\mathcal{K}}(\alpha)$ *of a formula* $\alpha \in \mathcal{L}$ *is the smallest index* i *such that* $\overline{\mathcal{E}_i^{\mathcal{K}}} \models \neg\alpha$. *If there is no such* i*, then let* $\mathrm{br}_{\mathcal{K}}(\alpha) = \infty$. *This is distinguished from the case of* $\mathrm{br}_{\mathcal{K}}(\alpha) = n$ *where* $\mathcal{E}_\infty^{\mathcal{K}}$ *is the first* $\mathcal{E}_i^{\mathcal{K}}$ *having* $\overline{\mathcal{E}_i^{\mathcal{K}}} \models \neg\alpha$.

We also introduce the following shorthand:

Definition 4. *For a knowledge base* $\mathcal{K}$ *and formula* α*, let* $\mathcal{E}_\alpha^{\mathcal{K}} = \mathcal{E}_r^{\mathcal{K}}$ *where* $r = \mathrm{br}_{\mathcal{K}}(\alpha)$. *The cases of* $r = \infty$ *and* $r = n$ *both correspond to* $\mathcal{E}_\alpha^{\mathcal{K}} = \mathcal{E}_\infty^{\mathcal{K}}$.

We note then that Rational Closure entailment $\approx_{\mathrm{RC}}$ can alternatively be expressed as follows:

Proposition 1. *For a knowledge base* $\mathcal{K}$ *and an entailment query* $\alpha \mathrel{|\!\sim} \beta$,

$$\mathcal{K} \approx_{\mathrm{RC}} \alpha \mathrel{|\!\sim} \beta \text{ iff } \mathrm{br}_{\mathcal{K}}(\alpha) = \infty \text{ or } \overline{\mathcal{E}_\alpha^{\mathcal{K}}} \models \alpha \to \beta.$$

3 Weak Justification

One of the main works of interest here is that of Chama [8] which proposes a notion of defeasible justification for Rational Closure according to an algorithm closely connected to the Rational Closure reasoning algorithm. The insight here is that we should follow the same process to eliminate more general statements, and once we have done so, to use classical tools to reason about the knowledge base—only in this case we obtain classical justifications instead of testing for classical entailment. We refer to these justifications as *weak justifications* to distinguish from classical justifications and the strong justifications we discuss later. We express this result for KLM propositional logic in the following definition (see Appendix A for the corresponding algorithm):

Definition 5. *A knowledge base* $\mathcal{J}$ *is a weak justification for a Rational Closure entailment* $\mathcal{K} \approx_{\mathrm{RC}} \alpha \mathrel{|\!\sim} \beta$ *if* $\overline{\mathcal{J}}$ *is a classical justification for* $\overline{\mathcal{E}_\alpha^{\mathcal{K}}} \models \alpha \to \beta$. *The set of weak justifications for* $\mathcal{K} \approx_{\mathrm{RC}} \alpha \mathrel{|\!\sim} \beta$ *is denoted* $\mathcal{J}_W\big(\mathcal{K}, \alpha \mathrel{|\!\sim} \beta\big)$.

3.1 Relevant Closure

Casini et al. [4] propose Relevant Closure which adapts the reasoning algorithm for Rational Closure so that we only retract the statements in a less specific rank that actually disagree with more specific statements in higher ranks with respect to the antecedent of the query. Relevant Closure is not rational; it does not obey all of the axioms of rational defeasible entailment.

For the sake of brevity, we do describe the reasoning algorithm procedurally. However, the essence of the Relevant Closure reasoning algorithm can be expressed simply using the following three definitions:

Definition 6. *A knowledge base* $\mathcal{J}$ *is an* ε*-justification for* $(\mathcal{K}, \alpha)$ *if* $\overline{\mathcal{J}}$ *is a classical justification for* $\overline{\mathcal{K}} \models \neg\alpha$.

Definition 7. *A statement* $\alpha \mid\sim \beta \in \mathcal{K}$ *is* relevant *for* $(\mathcal{K}, \gamma)$ *if* $a \mid\sim \beta$ *is an element of some* ε*-justification for* $(\mathcal{K}, \gamma)$*. Let* $R(\mathcal{K}, \gamma)$ *be the set of statements in* $\mathcal{K}$ *that are relevant for* $(\mathcal{K}, \gamma)$ *and* $R^{-}(\mathcal{K}, \gamma)$ *the set of statements in* $\mathcal{K}$ *not relevant to* $(\mathcal{K}, \gamma)$*.*

Definition 8. *We have* $\mathcal{K} \approx_{\text{RelC}} \alpha \mid\sim \beta$ *if* $\overline{\mathcal{E}_{\alpha}^{\mathcal{K}} \cup R^{-}(\mathcal{K}, \alpha)} \models \alpha \rightarrow \beta$*.*

What we have described here is Basic Relevant Closure, but Minimal Relevant Closure—the other definition of Relevant Closure entailment [4]—is based on a slightly altered version of relevance and the difference is not important for our purposes (i.e. minimal and basic relevance can be 'swapped out' by having $R(\cdot, \cdot)$ and $R^{-}(\cdot, \cdot)$ correspond to the form of relevance at hand).

Weak Justifications for Relevant Closure We identify the following analogue of weak justification for the case of Relevant Closure entailment:

Definition 9. *A knowledge base* $\mathcal{J}$ *is a weak justification for an entailment* $\mathcal{K} \approx_{\text{RelC}} \alpha \mid\sim \beta$ *if* $\overline{\mathcal{J}}$ *is a classical justification for* $\overline{\mathcal{E}_{\alpha}^{\mathcal{K}} \cup R^{-}(\mathcal{K}, \alpha)}$*.*

In other words, we ensure that the statements in the knowledge base considered not relevant to the query remain under our consideration when materialising just as they are when evaluating Relevant Closure entailment queries. We give a corresponding algorithm in Appendix A by adapting `WeakJustifyRC`.

3.2 Lexicographic Closure

Lexicographic Closure is another rational definition for defeasible entailment proposed by Lehmann [12] which is more permissive than Rational Closure. Like Rational Closure, Lexicographic Closure can be defined both semantically and algorithmically; we once again present the algorithmic definition. Lexicographic Closure can be seen as a refinement of Rational Closure, where we remove single statements instead of entire levels when inconsistencies arise during the reasoning process. We utilize the algorithm presented by Morris et al. [15] for propositional logic as the algorithm for Lexicographic Closure. However, while Morris et al. refine the ranking of statements, we instead refine the removal of statements as shown in `LexicographicClosure` in Algorithm 3.

Essentially, instead of removing an entire level R_i, this refinement weakens R_i by simultaneously considering all the ways of removing j statements from the level.

Weak Justifications for Lexicographic Closure The refined method presented in `LexicographicClosure` is equivalent to considering a series of sub-knowledge bases which are derived by replacing R_i with all the possible subsets R_i of size $m-j$, where m is the number of statements in R_i. Using this approach, the final entailment holds if the classical entailment holds in all the final sub-knowledge bases. Thus to provide a weak justification for the entailment, we can

Algorithm 3: LexicographicClosure

Input: A knowledge base $\mathcal{K}$ and DI $\alpha \mathrel{|\!\sim} \beta$
Output: **true**, if $\mathcal{K} \approx_{LC} \alpha \mathrel{|\!\sim} \beta$, otherwise **false**

$(R_0, ..., R_{n-1}, R_\infty, \text{n}) := \text{BaseRank}(\mathcal{K})$;
i := 0;
$R := \bigcup_{j=0}^{j<n} R_j$;
while $R_\infty \cup R \models \neg\alpha$ ***and*** $R \neq \emptyset$ **do**
  $R := R \setminus R_i$;
  m := $|R_i| - 1$;
  $R_{i,m} := \bigvee_{X \in \text{Subsets}(R_i, m)} \bigwedge_{x \in X} x$;
  while $R_\infty \cup R \cup \{R_{i,m}\} \models \neg\alpha$ ***and*** $m > 0$ **do**
    m := m-1;
    $R_{i,m} := \bigvee_{X \in \text{Subsets}(R_i, m)} \bigwedge_{x \in X} x$;
  end
  $R := R \cup \{R_{i,m}\}$;
  i := i+1;
end
return $R_\infty \cup R \models \alpha \rightarrow \beta$;

compute the classical justifications in the sub-knowledge bases and present these as the final justification. However, we wish to maintain a structure that refers to which justifications are responsible for the entailment of the statement in each sub-knowledge base. We therefore present a tuple as our final weak justification, where the i'th element of the tuple is a justification for the entailment of the statement in the i'th sub-knowledge base. This allows us to refer to individual statements in our knowledge base instead of a single combined formula.

To compute these weak justifications, we first modify `LexicographicClosure` to create `LexicographicClosureForJustifications`. This modification is performed by returning the variable m, which represents the subset size used in the final entailment computation, and the variable i, which represents the lowest complete level used in the final entailment computation, as an ordered pair (i, m) along with the final entailment result. We can define an algorithm `WeakJustificationsLC` that then uses this information to reconstruct the appropriate sub-knowledge bases and compute the justifications for each sub-knowledge base using classical methods as was done for Rational Closure. Taking the cross product of these justifications then yields the final set of tuples, which are the weak justifications. Full details of both algorithms are given in Appendix A. We demonstrate how weak justifications are computed for Lexicographic Closure in the following example.

Example 2. Suppose one has the knowledge base in Example 1 and consider once again the entailment of the statement 'penguins typically have eyes' ($p \mathrel{|\!\sim} e$). `LexicographicClosureForJustifications` returns **true**, along with $(1, 2)$. Thus we reconstruct the sub-knowledge bases $\mathcal{K}_1 = \{b \mathrel{|\!\sim} f, b \mathrel{|\!\sim} e\} \cup R_1 \cup R_\infty$, $\mathcal{K}_2 = \{b \mathrel{|\!\sim} f, b \mathrel{|\!\sim} s\} \cup R_1 \cup R_\infty$ and $\mathcal{K}_3 = \{b \mathrel{|\!\sim} e, b \mathrel{|\!\sim} s\} \cup R_1 \cup R_\infty$. The set of all justifications for $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ are $\mathcal{J}_1 = \{j_1, j_2\}$, $\mathcal{J}_2 = \{j_1\}$ and $\mathcal{J}_3 = \{j_2\}$ respectively, where:$j_1 = \{b \mid\sim f, p \mid\sim \neg f, p \rightarrow b\}$, $j_2 = \{b \mid\sim e, p \rightarrow b\}$. Thus our weak justifications will be the elements of the cross products of these sets. For example, (j_1, j_1, j_2) will be a weak justification for the entailment.

4 Strong Justification for Rational Closure

While weak justifications present an intuitive and simple approach to defining defeasible explanation, their level of description is arguably limited: we are only presenting information as to why the final classical entailment holds and are thus disregarding the rest of the reasoning process such as the determination of the base rank. In this section, we apply an intuitively more comprehensive definition for defeasible explanation, referred to as strong explanation, proposed by Brewka and Ulbricht [3] to KLM style reasoning to produce what we refer to as *strong justifications*. In particular, we look at defining what constitutes a strong justification for Rational Closure and explore an algorithm for computing these strong justifications. Note that the proofs of propositions and full details for all algorithms not presented in this section are given in Appendix B.

4.1 Overview of the Approach

A strong justification is a set $\mathcal{S}$ such that for any $\mathcal{S}'$ with $\mathcal{S} \subseteq \mathcal{S}' \subseteq \mathcal{K}$, $\mathcal{S}' \approx \alpha \mid\sim \beta$ and the previous statement does not hold for any $\mathcal{S}'' \subset \mathcal{S}$. This is essentially an extension of the definition for classical justification. The intuition here is that the strong justification contains 'just enough' of the statements such that the defeasible entailment always holds even if arbitrary statements from the knowledge base are added to the justification. Note that while weak justifications for Rational Closure are a subset of the knowledge base $\mathcal{K}$ that entails the final statement—and hence always obey at least one aspect of this criterion—they are not necessarily strong justifications. As an illustration of this, consider the following example:

Example 3. Suppose one has a knowledge base $\mathcal{K}$ containing the following information:

1. If something walks, it typically does not fly ($w \mid\sim \neg f$)
2. If something walks, it typically has legs ($w \mid\sim l$)
3. Pigeons typically fly ($p \mid\sim f$)
4. Pigeons typically walk ($p \mid\sim w$)

This knowledge base has the ranking shown in Figure 2a. Consider the entailment of the statement 'if something is a pigeon and it walks then it typically flies' ($p \wedge w \mid\sim f$). The weak justification for this statement is $\mathcal{W} = \{p \mid\sim f\}$.

Now consider what happens when the statement $w \mid\sim \neg f$ is added to $\mathcal{W}$. Ranking $\mathcal{W} \cup \{w \mid\sim \neg f\}$ yields the ranking shown in Figure 2b. However, now when computing the entailment of $p \wedge w \mid\sim f$, $R_0 \models \neg(p \wedge w)$ and so R_0 is removed. But then $\emptyset \models (p \wedge w) \rightarrow f$ is computed as the final entailment result, which does not hold. Thus $\mathcal{W}$ is not a strong justification since $\mathcal{W} \cup \{w \mid\sim \neg f\}$ does not entail our statement.

0	$w \mid\sim \neg f, w \mid\sim l$
1	$p \mid\sim w, p \mid\sim f$
∞	

(a) Initial ranking for $\mathcal{K}$

0	$w \mid\sim \neg f, p \mid\sim f$
∞	

(b) Ranking for $\mathcal{W} \cup \{w \mid\sim \neg f\}$

Fig. 2: Base ranking of statements for Example 3

The issue that arises in Example 3 is that we can add statements to our set that allow us to entail the negation the antecedent α off our entailed statement but not the negation of all the antecedents of our statements in our weak justification. This leads to some of the statements in our weak justification getting 'mixed into' lower ranks in which we can still 'disprove' α, which are then removed during the algorithm. However, we can extend $\mathcal{W}$ to create the strong justification $\mathcal{S} = \{p \mid\sim w, p \mid\sim f\}$. Here we ensure that the weak justification is always pushed above the ranks in which we can 'disprove' p, which are the ranks that contain justifications for $\neg p$.

4.2 Algorithm

We wish to define a procedure for extending weak justifications to form strong justifications as we did for Example 3. First we define the ranking of a set $\mathcal{K}' \subseteq \mathcal{K}$ in $\mathcal{K}$ to be the rank of the statement in $\mathcal{K}'$ with the lowest ranking in $\mathcal{K}$. For example, in Figure 2a, the set $\mathcal{K}' = \{w \mid\sim \neg f, p \mid\sim w\}$ has rank 0 in $\mathcal{K}$. We wish to ensure our weak justification is always ranked above the justifications for $\neg\alpha$, which we denote $\mathcal{J}_{\neg\alpha}$. We restrict ourselves to considering strong justifications for entailments where all justifications $\mathcal{J}_{\neg\alpha}$ and weak justifications have finite rank.

We start by considering every possible way of ranking the justifications $\mathcal{J}_{\neg\alpha}$ and then consider all possible ways of ranking weak justifications $\mathcal{W}$ such that they are one rank higher than the highest ranked $\mathcal{J}_{\neg\alpha}$ in $\mathcal{K}$. To do this we build up a sequences of subsets $(K_0, K_1, \ldots, K_x)$ where:

1. $br_{K_i}(\gamma) = i$
2. For all $K' \subset K_i$, $br_{K'}(\gamma) < br_{K_i}(\gamma)$

If a statement γ has rank n, that the justification for $\neg\gamma$ has rank $n-1$. Thus we use the statement $\gamma = \alpha$ for all justifications $\mathcal{J}_{\neg\alpha}$ and $\gamma = \alpha \wedge \neg\beta$ for all weak justifications $\mathcal{W}$. We use this approach instead of defining a fixed initial set to ensure the subsets in the sequence constructed are minimal. `ComputeSubsetSequences` shown in Algorithm 4, with the sub-process `Sequences` shown in Algorithm 5, computes the appropriate sequences for weak justifications. Proposition 2 states that `ComputeSubsetSequences` will always compute at least one sequence for some weak justification.

Proposition 2. *Let $\mathcal{K}$ be a knowledge base and γ a formula. Provided $br_{\mathcal{K}}(\gamma) \neq \infty$, `ComputeSubsetSequences` returns at least one sequence of sets.*

Algorithm 4: ComputeSubsetSequences

Input: A knowledge base $\mathcal{K}$, a formula α, the rank n $\leq br_{\mathcal{K}}(\alpha)$ required for α in the final subset of a sequence

Output: Set of all sequences $(K_1, K_2, ..., K_n)$, where α has $br_{K_i}(\alpha) = i$, and each K_i is minimal

i := 0;
$E_0 := \overline{\mathcal{K}}$;
repeat
    $E_{i+1} := \{\alpha \rightarrow \beta \in E_i \mid E_i \models \neg\alpha\}$;
    i := i+1;
until $E_{i-1} = E_i$;
$E_\infty := E_{i-1}$;
sequences := *Sequences*$((E_0, \ldots, E_\infty), \alpha, (\emptyset), n, 1)$;
return *sequences*;

Note in `Sequences`, *MinimalExtension*$(\alpha,\ A, B)$ computes the set of all sets M where M is attained by adding the minimum number of statements from A to B so that B entails α and *Minimize*$(A,\ \alpha)$ returns true if we can remove statements from A and maintain the ranking of α in A.

Algorithm 5: Sequences

Input: A knowledge base $\mathcal{K} = (E_0, E_1, ...E_\infty)$, a formula α, the current sequence of subsets $(K_0, \ldots, K_{i-1})$, the rank n of α in the final subset of the sequence, the index i of the current subset

Output: A set of sequences of length n

if $i > n$ **then**
    return $\{(K_0, \ldots, K_{i-1})\}$;
$\mathcal{A} := \neg\alpha \wedge (\bigwedge_{\beta \in Antecedents(K_{i-1})} \neg\beta)$;
S := *MinimalExtension*$(\mathcal{A}, E_{n-i}, K_{i-1})$;
$\mathcal{F} := \emptyset$;
if $S \neq \emptyset$ **then**
    for K_i *in* S **do**
        if $br_{K_i}(\alpha) = i$ *and Minimize*(K_i, α) *is* ***False*** **then**
            $\mathcal{F} := \mathcal{F} \cup Sequences(\mathcal{K}, \alpha, (K_0, \ldots, K_{i-1}, K_i), n, i+1)$;
        end
    end
return $\mathcal{F}$;

As a demonstration of how `ComputeSubsetSequences` creates a sequence, consider the following example.

Example 4. Consider the knowledge base and query in Example 3. `ComputeSubsetSequences` computes a single sequence (K_0, K_1, K_2) for the weak justifications, with $K_0 = \emptyset, K_1 = \mathcal{W} = \{p \mathrel{|\!\sim} w\}, K_2 = K_1 \cup \{p \mathrel{|\!\sim} w, w \mathrel{|\!\sim} \neg f\}$.

We define similar algorithms `ComputeGeneralSubsetSequences` and `GeneralSequences` for computing the sequences for all justifications $\mathcal{J}_{\neg\alpha}$. However, since we do not require these justifications to reach a certain rank, we do not need to work from E subsets. Instead statements are added from $\overline{\mathcal{K}}$. We also return a sequence when the final set can no longer be minimally extended, instead of requiring the algorithm to iterate a fixed number of times. Proposition 3 states the soundness and completeness of `ComputeSubsetSequences` for computing the minimal ways of ranking a statement.

Proposition 3. *Let $\mathcal{K}$ be a knowledge base and γ a formula. Provided $br_{\mathcal{K}}(\gamma) \neq \infty$, `ComputeGeneralSubsetSequences` computes exactly $M \subseteq \mathcal{K}$ such that for all $M' \subset M, br_{M'}(\gamma) < br_M(\gamma)$ i.e. M is minimal in terms of the ranking of γ.*

Given the ability to compute these sequences, we now define the formal extension of a weak justification to create a strong justification. If we choose a sequence $S_{\mathcal{W}}$ for some weak justification, we can consider all sequences for justifications $\mathcal{J}_{\neg\alpha}$ and use $S_{\mathcal{W}}$ to add the minimum number of statements to $\mathcal{W}$ to ensure it is ranked above $\mathcal{J}_{\neg\alpha}$ in each subset in each sequence. We consider each sequence $\mathcal{J}$ produced by `ComputeGeneralSubsetSequences` individually and compute all such minimal sets S for the sequence. To do this we define the algorithm `StrongSequences`. We start with S containing our weak justification and then iterate through each subset $K_i^{\mathcal{J}}$ in $\mathcal{J}$, checking whether $K_i^{\mathcal{J}} \cup S \models \neg m$ for all $m \mathrel{|\!\sim} n \in S$. If this does not hold, we use $K_{i+1}^{\mathcal{W}}$ from our weak justification sequence and consider all the ways of adding the minimum number of required statements to S so that $K_i^{\mathcal{J}} \cup S \models \neg m$ for all $m \mathrel{|\!\sim} n \in S'$, where S' is the previous iteration of S in a similar manner to Algorithm 5. We repeat this process until we have considered all sets in the sequence. Essentially, what we have done is create all the minimal sets S that ensure that whenever $\mathcal{J}_{\neg\alpha}$ has rank j in our sequence, our weak justifications has at least rank $j + 1$.

We then consider all ways of taking a minimal set for each sequence and taking the union of these sets. The smallest set created in this manner is then taken as our strong justification. This full process is defined by the algorithm `StrongJustification`. Proposition 4 states the correctness of this algorithm.

Proposition 4. *Let $\mathcal{K}$ be a knowledge base and $\alpha \mathrel{|\!\sim} \beta$ a defeasible implication. Provided $br_{\mathcal{K}}(\alpha) \neq \infty$, `StrongJustification` returns a strong justification for the entailment $\mathcal{K} \mathrel{|\!\approx}_{RC} \alpha \mathrel{|\!\sim} \beta$.*

We provide a simple demonstration of `StrongJustification` using our example from the introduction.

Example 5. Consider the knowledge base and query in Example 3. `ComputeSubsetSequences` produces the sequence in Example 4. `ComputeGeneralSubsetSequences` produces the single sequence $\mathcal{J} = (K_0, K_1)$

with $K_0^{\mathcal{J}} = \emptyset, K_1^{\mathcal{J}} = \{p \mid\sim \neg f, b \mid\sim f\}$. We start with $S = \mathcal{W} = \{p \mid\sim f\}$ for this sequence. We then consider whether $K_1^{\mathcal{J}} \cup S \models \neg p$. Since this does not hold, we add the statement $p \mid\sim w \in K_2$ to S. Since there are no more subsets in $\mathcal{J}$, we return $S = \{p \mid\sim \neg f, p \mid\sim w\}$ as the single minimal set and, since there are no other sequences, S is returned as our final strong justification.

4.3 Limitations for the Approach

While we have defined strong justification as an extension of weak justifications, not every weak justification can be extended to create a strong justification. As an illustration of this, consider the following example.

Example 6. Suppose one has the knowledge base $\mathcal{K}$ shown in Figure 3 and consider the entailment of the statement $sp \wedge p \mid\sim m \vee w$. This entailment has two weak justifications:

1. $\mathcal{W}_1 = \{sp \mid\sim f, f \mid\sim m\}$
2. $\mathcal{W}_2 = \{sp \mid\sim f, f \mid\sim w\}$

$\mathcal{W}_1$ cannot be extended to create a strong justification. If initially $\mathcal{S} = \mathcal{W}_1$, then to ensure $\mathcal{S} \cup \{p \mid\sim \neg f\} \mid\approx_{RC} sp \wedge p \mid\sim m \vee w$, the statements $\{l \mid\sim \neg x, f \mid\sim w, w \mid\sim x, w \rightarrow l\}$ need to be added to $\mathcal{S}$. Then $\mathcal{S} = \{l \mid\sim \neg x, f \mid\sim w, w \mid\sim x, w \rightarrow l, sp \mid\sim f, f \mid\sim m\}$. But now if $f \mid\sim m$ is removed from $\mathcal{S}$, we still have $\mathcal{S}' \mid\approx sp \wedge p \mid\sim m \vee w$ for all $\mathcal{S}' \subseteq \mathcal{S} \subseteq \mathcal{K}$. Thus due to the minimality property of strong justifications, $\mathcal{W}_1$ cannot be extended to create a strong justification.

0	$p \mid\sim \neg f, l \mid\sim \neg x$
1	$sp \mid\sim f, f \mid\sim m, f \mid\sim w, w \mid\sim x$
∞	$w \rightarrow l$

Fig. 3: Base ranking for statements in Example 6

It is also not the case that all strong justifications can be defined as an extension of a weak justification. As a demonstration of this fact, consider the following example.

Example 7. Suppose one has a knowledge base $\mathcal{K}$ shown in Figure 4 and consider the entailment of the statement $p \mid\sim s$. This has the weak justification $\mathcal{W} = \{p \mid\sim \neg f, \neg f \mid\sim s\}$. However, consider rather $\mathcal{B} = \{b \mid\sim s, p \rightarrow b\}$ as the base set for extension. $\mathcal{B}$ is not a weak justification since it contains information removed during the Rational Closure algorithm. However, $\mathcal{B}$ can be extended to form the strong justification

$$\mathcal{S} = \{b \mid\sim s, \neg f \mid\sim s, r \mid\sim w, \neg f \mid\sim \neg w, \neg f \mid\sim r, p \rightarrow b\}.$$

First notice $\mathcal{S} \mid\approx_{RC} p \mid\sim s$. Now consider what happens if any statements are added to $\mathcal{S}$. If we add any statements that cause $\mathcal{B}$ to be thrown away, namely $p \mid\sim \neg f$, the set then contains $\mathcal{W}$ and so the entailment still holds.

0	$b \mid\sim f, b \mid\sim s, r \mid\sim w$
1	$p \mid\sim \neg f, \neg f \mid\sim \neg w, \neg f \mid\sim r, \neg f \mid\sim s$
∞	$p \to b$

Fig. 4: Base ranking of statements for Example 7

5 Properties of Weak Justification

Weak justification has currently only been explored in terms of reasoning algorithms (such constructions as the exceptionality sequence $\mathcal{E}$ and base ranks) and therefore an interesting question is whether it can be characterised in a more intuitive manner. In this section, we show that for every postulate of rationality there is a corresponding property obeyed by weak justification. For the sake of simplicity, our presentation here will be limited to the case of Rational Closure and therefore we assume for this section that $\approx$ refers to $\approx_{\mathrm{RC}}$. We begin by considering a strengthening of the postulates for rationality given by Lehmann and Magidor [13]:

1. Left logical equivalence *(LLE)*. If $\mathcal{K} \approx \alpha \leftrightarrow \beta$ and $\mathcal{K} \approx \alpha \mid\sim \gamma$ then $\mathcal{K} \approx \beta \mid\sim \gamma$.
2. Right weakening *(RW)*. If $\mathcal{K} \approx \alpha \to \beta$ and $\mathcal{K} \approx \gamma \mid\sim \alpha$ then $\mathcal{K} \approx \gamma \mid\sim \beta$.
3. *And*. If $\mathcal{K} \approx \alpha \mid\sim \beta$ and $\mathcal{K} \approx \alpha \mid\sim \gamma$ then $\mathcal{K} \approx \alpha \mid\sim \beta \land \gamma$.
4. *Or*. If $\mathcal{K} \approx \alpha \mid\sim \gamma$ and $\mathcal{K} \approx \beta \mid\sim \gamma$ then $\mathcal{K} \approx \alpha \lor \beta \mid\sim \gamma$.
5. Reflexivity *(Ref)*. $\mathcal{K} \approx \alpha \mid\sim \alpha$.
6. Cautious Monotonicity *(CM)*. If $\mathcal{K} \approx \alpha \mid\sim \gamma$ and $\mathcal{K} \approx \alpha \mid\sim \beta$ then $\mathcal{K} \approx \alpha \land \beta \mid\sim \gamma$.
7. Rational Monotonicity *(RM)*. If $\mathcal{K} \approx \alpha \mid\sim \gamma$ and $\mathcal{K} \not\approx \alpha \mid\sim \neg\beta$ then $\mathcal{K} \approx \alpha \land \beta \mid\sim \gamma$.

This is a strengthening of the KLM postulates because *LLE* has the condition $\mathcal{K} \approx \alpha \leftrightarrow \beta$ in favour of $\alpha \equiv \beta$ and *RW* has $\mathcal{K} \approx \alpha \to \beta$ in favour of $\alpha \models \beta$. (Note here that $\mathcal{K} \approx \alpha \leftrightarrow \beta$ is for example is a shorthand for $\mathcal{K} \approx \neg(\alpha \leftrightarrow \beta) \mid\sim \bot$ as discussed in the background section.) This provides a more useful perspective for our purposes.

Our approach is to consider how defeasible justification applies to each of these axioms. Take for instance the example of *And*. The insight here is that if α typically implies β, and α typically implies γ, then not only should we be able to conclude that α typically implies β and γ, we should be able to conclude it *by the same token*. In this section we formalise this idea, and a similar idea for each postulate above, in relation to weak justification. The following concept helps us state these results:

Definition 10. *A knowledge base $\mathcal{D} \subseteq \mathcal{K}$ is deciding for an entailment $\mathcal{K} \approx \alpha \mid\sim \beta$ if $\mathcal{D} \subseteq \mathcal{E}^{\mathcal{K}}_{\alpha}$ and $\overline{\mathcal{D}} \models \alpha \to \beta$.*

For a given entailment, any deciding knowledge base is always a superset of a weak justification and all weak justifications are deciding (refer to Definition 5). We also have the following results for deciding knowledge bases:

Proposition 5. *If $\mathcal{D}$ is a deciding knowledge base for an entailment $\mathcal{K} \approx \alpha \mid\sim \beta$ and $\mathrm{br}_{\mathcal{K}}(\alpha) \neq \infty$, then $\mathcal{D} \approx \alpha \mid\sim \beta$.*

Proposition 6. *If $\mathcal{D}$ is a deciding knowledge base for an entailment $\mathcal{K} \approx \alpha \mid\sim \beta$ and we have $\mathcal{D} \approx \alpha \mid\sim \beta$, then $\mathcal{J}_W\left(\mathcal{D}, \alpha \mid\sim \beta\right) \subseteq \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \beta\right)$.*

Proofs for these results, as well as all results in this section, are given in Appendix C. We can now state a result corresponding to each axiom of rational defeasible entailment:

Theorem 1. *For any knowledge bases $\mathcal{K}, \mathcal{J}_1, \mathcal{J}_2$,*

(LLE) *if $\mathcal{J}_1 \in \mathcal{J}_W(\mathcal{K}, \alpha \leftrightarrow \beta)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \gamma\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \beta \mid\sim \gamma$;*

(RW) *if $\mathcal{J}_1 \in \mathcal{J}_W(\mathcal{K}, \alpha \rightarrow \beta)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \gamma \mid\sim \alpha\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \gamma \mid\sim \beta$;*

(And) *if $\mathcal{J}_1 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \beta\right)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \gamma\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \alpha \mid\sim \beta \wedge \gamma$;*

(Or) *if $\mathcal{J}_1 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \gamma\right)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \beta \mid\sim \gamma\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \alpha \vee \beta \mid\sim \gamma$.*

(Ref) *$\mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \alpha\right) = \{\emptyset\}$.*

(CM) *if $\mathcal{K} \approx \alpha \mid\sim \gamma$ and $\mathcal{K} \approx \alpha \mid\sim \beta$, every $\mathcal{J} \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \gamma\right)$ is deciding for $\mathcal{K} \approx \alpha \wedge \beta \mid\sim \gamma$;*

(RM) *if $\mathcal{K} \approx \alpha \mid\sim \gamma$ and $\mathcal{K} \not\approx \alpha \mid\sim \neg\beta$, every $\mathcal{J} \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mid\sim \gamma\right)$ is deciding for $\mathcal{K} \approx \alpha \wedge \beta \mid\sim \gamma$.*

6 Conclusions & Future Work

We have extended the principle of weak justification, previously only explored for Rational Closure, to the case of Relevant and Lexicographic Closure and proposed algorithms for enumerating these justifications. We then evaluated a revised definition of strong justification in light of certain issues and proposed an algorithm that enumerates these justifications for Rational Closure. This is, to our knowledge, the first application of strong explanation to KLM and may offer an alternative to weak justification that is perhaps more comprehensive as far as the resulting justifications are concerned. Our final result is a characterisation of weak justification in relation to the KLM postulates for rationality. This provides evidence that weak justification is a sound and generalisable notion of justification for KLM-style defeasible entailment and illustrates similarities between weak justification for the defeasible case and classical justification for the classical case.

Since the algorithm we propose for enumerating strong justifications and the declarative characterisation of weak justification were limited to the case of Rational Closure, further work might seek to apply this result to other notions of defeasible entailment for KLM or perhaps generally to rational defeasible entailment. Another possibility is to consider how these ideas apply to KLM description logic [5, 7, 14].

A Weak Justifications

Note we assume that `ComputeAllJustifications` is some algorithm that accepts a classical knowledge base $\mathcal{K}$ and an entailment query $\alpha \in \mathcal{L}$ and returns the set of all justifications for $\mathcal{K} \models \alpha$ (or the empty set if the entailment does not hold) as described by Horridge [10].

Algorithm 6: WeakJustifyRC

Data: A knowledge base $\mathcal{K}$ and a query $\alpha \mid\!\sim \beta$

Result: The set of weak justifications for $\mathcal{K} \approx_{\text{RC}} \alpha \mid\!\sim \beta$

$(R_0, R_1, ..., R_\infty,$ n) := BaseRank(K);
i := 0;
$R := \bigcup_{i=0}^{j<n} R_j$;
while $R_\infty \cup R \models \neg\alpha$ ***and*** $R \neq \emptyset$ **do**
    $R := R \setminus R_i$;
    i := i+1;
end
return $\big\{\{\gamma \mid\!\sim \delta \mid \gamma \to \delta \in \mathcal{J}\} \mid \mathcal{J} \in \text{ComputeAllJustifications}\,(R, A \to B)\big\}$;

A.1 Relevant Closure

Algorithm 7: WeakJustifyRelC

Data: A knowledge base $\mathcal{K}$, a query $\alpha \mid\!\sim \beta$ and the value of $R^-(\mathcal{K}, \alpha)$ which we denote as M

Result: The set of weak justifications for $\mathcal{K} \approx_{\text{RelC}} \alpha \mid\!\sim \beta$

$(R_0, R_1, ..., R_\infty,$ n) := BaseRank(K);
i := 0;
$R := \bigcup_{i=0}^{j<n} R_j$;
while $R_\infty \cup R \models \neg\alpha$ ***and*** $R \neq \emptyset$ **do**
    $R := R \setminus R_i$;
    i := i+1;
end
return
$\big\{\{\gamma \mid\!\sim \delta \mid \gamma \to \delta \in \mathcal{J}\} \mid \mathcal{J} \in \text{ComputeAllJustifications}\,(R \cup M, A \to B)\big\}$;

A.2 Lexicographic Closure

Algorithm 8: LexicographicClosureForJustifications

Input: A knowledge base $\mathcal{K}$ and DI $\alpha \mid\!\sim \beta$

Output: **true**, if $K \approx_{LC} \alpha \mid\!\sim \beta$, otherwise **false**, and an ordered pair (i, m) describing a rank and subset size

$(R_0, ..., R_{n-1}, R_\infty,$ n) := BaseRank($\mathcal{K}$);
i := 0;
m := 0;
$R := \bigcup_{j=0}^{j<n} R_j$;
while $R_\infty \cup R \models \neg\alpha$ ***and*** $R \neq \emptyset$ **do**
    $R := R \setminus R_i$;
    m := $|R_i|$-1;
    $R_{i,m} := \bigvee_{X \in \text{Subsets}(R_i, m)} \bigwedge_{x \in X} x$;
    while $R_\infty \cup R \cup \{R_{i,m}\} \models \neg\alpha$ ***and*** $m > 0$ **do**
        m := m-1;
        $R_{i,m} := \bigvee_{X \in \text{Subsets}(R_i, m)} \bigwedge_{x \in X} x$;
    end
    $R := R \cup \{R_{i,m}\}$;
    i := i+1;
end
return $R_\infty \cup R \models \alpha \rightarrow \beta$, (i, m);

Algorithm 9: WeakJustificationsLC

Input: A knowledge base $\mathcal{K}$ and DI $\alpha \mid\!\sim \beta$

Output: Set of all justifications for $K \approx_{LC} \alpha \mid\!\sim \beta$

$(R_0, ..., R_{n-1}, R_\infty,$ n) := *BaseRank*($\mathcal{K}$);
(i, m) := *LexicographicClosureForJustifications*($\mathcal{K}$, $\alpha \mid\!\sim \beta$);
$R := \bigcup_{j=i}^{j<n} R_j$;
$\mathcal{J} := \emptyset$;
if $m > 0$ **then**
    subsets := *Subsets*(R_{i-1}, m);
    i := 0;
    for S *in subsets* **do**
        $\mathcal{F} := S \cup R \cup R_\infty$;
        $\mathcal{J}_i :=$ *ComputeAllJustifications*$(\mathcal{F}, \alpha \rightarrow \beta)$;
        i := i+1;
    end
    $\mathcal{J} := \mathcal{J}_0 \times \mathcal{J}_1 \times \ldots \times \mathcal{J}_{i-1}$;
else
    $\mathcal{J} :=$ *ComputeAllJustifications*$(R \cup R_\infty)$;
return $\mathcal{J}$;

B Strong Justifications for Rational Closure

Algorithm 10: ComputeGeneralSubsetSequences

Input: A knowledge base $\mathcal{K}$, a formula α

Output: Set of all sequences $(K_0, K_1, ..., K_n)$, where α has $br_{K_i}(\alpha) = i$, and each K_i is minimal

sequences := $\mathit{GeneralSequences}(\overline{\mathcal{K}}, \alpha, (\emptyset), 1)$;

return *sequences*;

Algorithm 11: GeneralSequences

Input: A materialized knowledge base $\mathcal{K}$, a formula α, the current sequence of subsets $(K_0, \ldots, K_{i-1})$, the index i of the current subset

Output: A set of sequences of varying lengths

$\mathcal{A} := \neg\alpha \wedge (\bigwedge_{\beta \in Antecedents(K_{i-1})} \neg\beta)$;
S := $\mathit{MinimalExtension}(\mathcal{A}, \mathcal{K}, K_{i-1})$;
$\mathcal{F} := \emptyset$;
if $S \neq \emptyset$ **then**
  for K_i *in* S **do**
    if $br_{K_i}(\alpha) = i$ *and Minimize(K_i, α) is* ***False*** **then**
      $\mathcal{F} := \mathcal{F} \cup \mathit{GeneralSequences}(\mathcal{K}, \alpha, (K_0, \ldots, K_{i-1}, K_i), i+1)$;
    end
  end
if $\mathcal{F} = \emptyset$ **then**
  $\mathcal{F} := \{(K_0, \ldots, K_{i-1})\}$;
return $\mathcal{F}$;

Proposition 7. *Let $\mathcal{K}$ be a knowledge base, γ a formula and $M \subseteq \mathcal{K}$ such that for all $M' \subseteq M, br_{M'}(\gamma) < br_M(\gamma)$ i.e. M is minimal in terms of the ranking of γ. Let $i = br_{\mathcal{K}}(\gamma)$. Provided $i \neq \infty$, we can express M as a sequence of set $(K_0, K_1, \ldots, K_i)$ such that*

1. $br_{K_j}(\gamma) = j$
2. *For all* $K' \subset K_j, br_{K'}(\gamma) < br_{K_j}(\gamma)$

Proof. Let M, $\mathcal{K}$ and γ be as described in Proposition 7. We will construct a sequence of sets and prove that this sequence adheres to conditions (1) and (2).

Firstly $i = br_M(\gamma)$. We recursively construct a sequence of sets as follows:

- $M_0 = M$
- $M_{x+1} = \mathit{MinimizeSet}(M_x \setminus R_0^{M_x}, \gamma)$

MinimizeSet(A, γ) returns the smallest subset A' of A such that $br_{A'}(\gamma) = br_A(\gamma)$. If we set

$$K_0 = \emptyset, K_1 = M_{i-1}, \ldots, K_j = M_{i-j}, \ldots, K_i = M_{i-i} = M_0$$

we claim we have constructed an appropriate sequence. We will prove this is the case by first proving the statement '*For* M_y*,* M_y *is minimal and* $br_{M_y}(\gamma) = i - y$' using induction on y.

(BC) Let y = 0. Thus $M_0 = M$. By assumption M is minimal and $br_M(\gamma) = i = i - 0 = i - y$. Our statement is true for our base case.

(IH) Let $n \in \mathbb{N}$ with $0 \leq n < i$ and suppose our statements is true for M_n. We wish to show the statement is true for M_{n+1}. We derive M_{n+1} from M_n by removing all statements with rank 0 in M_n from M_n and minimizing the resulting set. It thus follows from the construction that M_{n+1} is minimal and so we must just prove that $br_{M_{n+1}}(\gamma) = i - (n + 1)$. Let

$$E_0^{M_n}, E_1^{M_n}, \ldots, E_{i-n-1}^{M_n}, E_{i-n}^{M_n}, \ldots, E_\infty^{M_n}$$

be the E sets for M_n. If we define $M' = M_n \setminus R_0^{M_n}$ we have $M' = E_1^{M_n}$. Thus

$$E_0^{M'} = E_1^{M_n}, E_1^{M'} = E_2^{M_n}, \ldots, E_{i-n-1}^{M'} = E_{i-n}^{M_n}, E_n^{M'} = E_{i-n+1}^{M_n}$$

Since $br_{M_n}(\alpha) = i - n$, we have that $E_{i-n}^{M_n}$ is the first E_x such that $E_x^{M_n} \not\models \neg\alpha$. Thus $E_{i-n-1}^{M'}$ is the first E_x for M' for which $E_x^{M'} \not\models \neg\alpha$ and so $br_{M'}(\alpha) = i - n - 1 = i - (n + 1)$. Finally note that $br_{M'}(\gamma) = br_{M_{n+1}}(\gamma)$ since $M_{n+1} = \mathit{MinimizeSet}(M')$. So we have $br_{M_{n+1}}(\gamma) = i - (n + 1)$ and thus are done.

Finally, since $br_{M_y}(\gamma) = i - y$ and $K_j = M_{i-j}$, $br_{K_j}(\gamma) = br_{M_{i-j}}(\gamma) = i - (i - j) = j$ and K_j is minimal. Thus our sequence satisfies both criteria and so we are done.

Proposition 8. *Let* $\mathcal{K}$ *be a knowledge base and* γ *a formula. Provided* $br_{\mathcal{K}}(\gamma) \neq \infty$*,* `ComputeGeneralSubsetSequences` *computes exactly* $M \subseteq \mathcal{K}$ *such that for all* $M' \subset M, br_{M'}(\gamma) < br_M(\gamma)$ *i.e. M is minimal in terms of the ranking of* γ*.*

Proof. Firstly note that since $\mathcal{K}$ is finite, *MinimalExtension* will eventually return an empty set and thus the algorithm will always terminate. To prove Proposition 8, we need to show that:

1. Every K_i in every sequence is minimal i.e. if $K' \subset K_i, br_{K'}(\gamma) < br_{K_i}(\gamma)$
2. Every $M \subseteq \mathcal{K}$ such that for all $M' \subseteq M, br_{M'}(\gamma) < br_M(\gamma)$ appears in a sequence.

The proof of (1) is trivial, since as part of the computation we check if each K_i is minimal before adding it to a sequence.

The proof of (2) follows from Proposition 7. We can express M as a sequence $(K_0, K_1, \ldots, K_i)$ by using the construction presented in the proof for Proposition 7. We need to show we can construct this sequence using the approach given in `GeneralSequences`. We have $K_0 = \emptyset$ by construction so trivially K_0 will be in a sequence. We will once again use induction to show that each K_j for $j > 0$ is a part of a sequence.

(BC) Let j = 1. Then since $br_{K_1}(\gamma) = 1$, $K_1 \subset \mathcal{K}$ such that $K_1 \models \neg\gamma$ and K_1 is minimal. Thus K_1 will be a part of S = $\mathit{MinimalExtension}(\mathcal{A} = \neg\gamma, \mathcal{K}, \emptyset)$

and since $br_{K_1}(\alpha) = 1$ and K_1 is minimal, K_1 will be added to a sequence $(K_0, K_1, \ldots)$.

(IH) Let $1 \leq j < i$ and suppose K_n is part of a sequence. We wish to show K_{n+1} will be added to a sequence. To do this we need to show:

(a) $K_{n+1} \setminus K_n$ is a minimal extension of K_n that allows us to entail $\mathcal{A} = \neg\gamma \wedge (\bigwedge_{\beta \in Antecedents(K_n)} \neg\beta)$
(b) K_{n+1} is minimal
(c) $br_{K_{n+1}}(\alpha) = n + 1$

For the proof of (a), note that $K_{n+1} = M_{i-n-1}$ and $K_n = M_{i-n}$. Since we get M_{i-n} by removing $R_0^{M_{i-n-1}}$ from M_{i-n-1} and minimizing, we have $M_{i-n} \subseteq E_1^{M_{i-n-1}}$. Thus $M_{i-n-1} \models \neg\beta_1 \wedge \neg\beta_2 \wedge \ldots$ for $\beta_x \in Antecedents(M_{i-n})$. Let $X = M_{i-n-1} \setminus M_{i-n}$. Suppose X is not minimal i.e. we can disprove all the antecedents in M_{i-n} using some subset X' of X. But then $br_{M_{i-n} \cup X'}(\gamma) \geq n + 1 = br_{M_{i-n-1}}(\gamma)$ with $M_{i-n} \cup X' \subset M_{i-n-1}$, which contradicts the minimality of M_{i-n-1}. Thus X must be minimal. Also note since $M_{i-n} \subseteq M_{i-n-1}$ and $M_{i-n} = K_n \models \neg\gamma$ since it is part of the sequence, we must have $M_{i-n-1} \models \neg\gamma$. So X will be returned by S = $MinimalExtension(\mathcal{A}, \mathcal{K}, K_n)$ when $\mathcal{A} = \neg\gamma \wedge (\bigwedge_{\beta \in Antecedents(K_n)} \neg\beta)$, allowing as to create K_{n+1}. Conditions (2) and (3) are satisfied based on our definition of the sequence in the proof for Proposition 7. So K_{n+1} will be added to a sequence and thus we are done.

Since each set in the sequence for M appears in a sequence, this means M itself appears in a sequence since $M = K_i$.

Proposition 9. *Let $\mathcal{K}$ be a knowledge base and γ a formula. Provided $br_{\mathcal{K}}(\gamma) \neq \infty$,* `ComputeSubsetSequences` *returns at least one sequence of sets.*

Proof. Let $\mathcal{K}$ and γ be as described in Proposition 9. Firstly note that since `Sequences` recurses to a maximum depth of n+1 for each sequence, the algorithm must always terminate. Now consider $M \subseteq \mathcal{K}$ such that $br_M(\gamma) = br_{\mathcal{K}}(\gamma) = j$ and for any $M' \subset M$, $br_{M'}(\gamma) < br_M(\gamma)$ and let $n \in \mathbb{N}$ with $n < j$. By Proposition 7, we can compute a sequence $(K_0, K_1, \ldots, K_n, \ldots, K_j)$ using the approach given in the proof. We claim if we take $(K_0, K_1, \ldots, K_n)$, we will have a sequence computed by `ComputeSubsetSequences`.

First we will prove that for each $M_s = K_{j-s}$, we have that $M_s \subseteq E_s^M$. We will prove this by induction on s.

(BC) Let s = 1. M_1 is constructed by taking $K' = M_0 \setminus R_0^{M_0} = M \setminus R_0^M$ and minimizing it. Thus $M_1 \subseteq K' \subseteq E_1^M$.

(IH) Let $1 \leq s < j$ and suppose $M_s \subseteq E_s^M$. We wish to show that $M_{s+1} \subseteq E_{s+1}^M$. We construct M_{s+1} by taking $K' = M_s \setminus R_0^{M_s}$ and minimizing it. Since $M_s \subseteq E_s^M$, $E_0^{M_s} \subseteq E_s^M$ and thus $E_1^{M_s} \subseteq E_{s+1}^M$. But $M_{s+1} \subseteq K' \subseteq E_1^{M_s} \subseteq E_{s+1}^M$ and so we are done.

Thus we have that $K_x = M_{j-x} \subseteq E_{j-x}^M$. Since $j > n$, we have $j - 1 \geq n$ and thus $j - x \geq n - (x - 1)$. So $K_x \subseteq E_{j-x}^M \subseteq E_{n-(x-1)}^M \subseteq E_{n-(x-1)}^{\mathcal{K}} \subseteq E_{n-x}^{\mathcal{K}}$. So K_x is contained in $E_{n-x}^{\mathcal{K}}$ for each x and thus can be used to create an extension for

K_{x-1}, meaning we are not limited by this restriction. So we can follow the same proof structure as the second half of the proof for Proposition 8 to prove that $(K_0, K_1, \ldots, K_n)$ is in fact computed and returned by `ComputeSubsetSequence`.

Algorithm 12: StrongSequences

Input: A sequence $\mathcal{N} = (S_0, S_1, \ldots, S_n)$ for a formula α, the formula α, a sequence $\mathcal{M} = (J_0, J_1, \ldots, J_m)$ for a formula β where $m < n$, the formula β, the index i of the current subset, the current minimal entailing set $\mathcal{S}$

Output: A set of sets $\mathcal{S}$ such that $br_{\mathcal{S} \cup J_i}(\alpha) > br_{\mathcal{S} \cup J_i}(\beta)$ for all $i \leq m$ and $\mathcal{S}$ is minimal

if $i > m$ **then**
  return $\{\mathcal{S}\}$;
$\mathcal{A} := \neg\alpha \wedge (\bigwedge_{\beta \in Antecedents(\mathcal{S})} \neg\beta)$;
$\mathcal{B} :=$ *MinimalExtension*$(\mathcal{A}, S_{i+(n-m)}, J_i \cup \mathcal{S})$;
$\mathcal{F} := \emptyset$;
for x *in* $\mathcal{B}$ **do**
  S' $= (x \setminus J_i) \cup \mathcal{S}$;
  if $br_x(\alpha) = br_x(\beta) + 1$ *and Minimize*(S, α, S') *is* ***False*** **then**
    $\mathcal{F} := \mathcal{F} \cup$ *StrongSequences*$(\mathcal{N}, \alpha, \mathcal{M}, i+1, S')$;
  end
end
$\mathcal{S} := \mathcal{F}$;
return $\mathcal{S}$;

Note that in the `StrongSequences` algorithm we use a different version of *Minimize*(A, α, B) where it returns true if we can remove statements $x \in B$ from A and maintain the same ranking for α in A.

Proposition 10. *Given a sequence $\mathcal{N} = (S_0, S_1, \ldots, S_n)$ for a formula γ where $br_{S_i} = i$, a formula γ, a sequence $\mathcal{M} = (J_0, J_1, \ldots, J_m)$ for a formula β where $br_{J_i}(\beta) = i$ and $m < n$, the formula β, the index 0 and the empty set as input,* `StrongSequences` *computes all minimal sets S such that:*

1. $br_{S \cup J_i}(\gamma) > br_{S \cup J_i}(\beta)$ *for all* $0 \leq i \leq m$.
2. *Condition (1) does not hold for any* $S' \subset S$.

Proposition 10 is yet to be proven.

Observation 1 *Giordano et al. [9]* $\mathcal{K} \approx_{RC} \alpha \mid\sim \beta$ *iff* $br_{\mathcal{K}}(\alpha) < br_{\mathcal{K}}(\alpha \wedge \neg\beta)$ *or* $br_{\mathcal{K}}(\alpha) = \infty$

Proposition 11. *Let $\mathcal{K}$ be a knowledge base and $\alpha \mid\sim \beta$ a defeasible implication. Provided $br_{\mathcal{K}}(\alpha) \neq \infty$,* `StrongJustification` *returns a strong justification for the entailment $\mathcal{K} \approx_{RC} \alpha \mid\sim \beta$.*

Algorithm 13: StrongJustification

Input: A knowledge base $\mathcal{K}$, the DI $\alpha \mathrel{|\!\sim} \beta$ for the entailment
Output: A strong justification $\mathcal{S}$

$\mathcal{J} := \mathit{ComputeGeneralSubsetSequences}(\mathcal{K}, \alpha)$;
m := $\max\{br_{\mathcal{K}}(j) : j \in \mathcal{J}\}$;
sequences = $\mathit{ComputeSubsetSequences}(\mathcal{K}, \alpha \wedge \neg\beta, m+1)$;
Choose K' = $(K_0, K_1, \ldots, K_{m+1})$ from sequences;
n := $|\mathcal{J}|$;
Let $S_1, S_2, \ldots, S_n$ be the set of minimal entailing sets associated with each sequence;
for J_i *in* $\mathcal{J}$ **do**
  $S_i := \mathit{StrongSequences}(K', \alpha \wedge \neg\beta, J_i, \beta, 0, \emptyset)$;
end
$\mathcal{B} = \{s_1 \cup s_2 \cup \ldots \cup s_n : s_i \in S_i\}$;
x = $\min\{|X| : X \in \mathcal{B}\}$;
Choose $\mathcal{S}$ from $\mathcal{B}$ such that $|\mathcal{S}| = x$;
return S;

Proof. Let $\mathcal{K}$ and $\alpha \mathrel{|\!\sim} \beta$ be as described in Proposition 11. Firstly note that if all the sub-processes of
`StrongJustification` terminate, so will `StrongJustification`. Also note that by Proposition 9, sequences is non-empty and thus we can always choose a K' to use for strengthening. Finally, note $\mathcal{J}$ will always be non-empty. In the case where $br_{\mathcal{K}}(\alpha) = 0$, it will just contain $\emptyset$. So `StrongJustification` will always return some set $\mathcal{S}$.

To show that $\mathcal{S}$ returned by `StrongJustification` is a strong justification, we must show that:

1. For any $\mathcal{S} \subseteq S' \subseteq K$, $S' \approx_{RC} \alpha \mathrel{|\!\sim} \beta$
2. For any $S'' \subset \mathcal{S}$, condition (1) does not hold.

For the proof of (1), suppose $\mathcal{S} \subseteq S' \subseteq K$ and let M be such that $br_M(\alpha) = br_{S'}(\alpha)$ and for all $M' \subset M, br_{M'}(\alpha) < br_M(\alpha)$. By Proposition 8, M is part of a sequence computed by `ComputeGeneralSubsetSequences`. Let $\mathcal{S}_M$ be the minimal entailing set for this sequence included in $\mathcal{S}$. So $\mathcal{S}_M \cup M \subseteq \mathcal{S} \cup M \subseteq S'$ and thus $br_{S'}(\alpha \wedge \neg\beta) \geq br_{\mathcal{S} \cup M}(\alpha \wedge \neg\beta) \geq br_{\mathcal{S}_M \cup M}(\alpha \wedge \neg\beta) > br_{\mathcal{S}_M \cup M}(\alpha)$. But $br_{\mathcal{S}_M \cup M}(\alpha) = br_M(\alpha) = br_{S'}(\alpha)$ since $S_M \cup M \subseteq S'$ and M is minimal. Thus $br_{S'}(\alpha \wedge \neg\beta) > br_{S'}(\alpha)$ and so $S' \approx_{RC} \alpha \mathrel{|\!\sim} \beta$.

For the proof of (2), let $S'' \subset \mathcal{S}$ and let $z \in \mathcal{S} \setminus S''$. Let $S_1, S_2, \ldots, S_m$ be the minimal entailing sets that contained x and $X_1, X_2, \ldots, X_n$ be their associated sequences. If $br_{x_i \cup S''}(\alpha \wedge \neg\beta) > br_{x_i \cup S''}(\alpha)$ for all x_i in each X_j, then there is some minimal subset J_j of S'' that ensures this condition for each j. But then, by Proposition 10, each of there J_j would have been computed for each sequence and thus $S'' \subseteq \mathcal{B}$. Since $S'' \subset \mathcal{S}$, $|S''| < |\mathcal{S}|$. But this contradicts the fact that $\mathcal{S}$ is chosen to be a set of minimum size in $\mathcal{B}$. Thus S' cannot exist and so $\mathcal{S}$ is in fact minimal.

C Properties of Weak Justification

As discussed in Section 5, we only consider the case of Rational Closure entailment, denoted here as $\approx$. We first introduce some lemmas which will help us with the proofs in this section:

Lemma 1. *If $\mathcal{K} \approx \alpha \rightarrow \beta$ for a knowledge base $\mathcal{K}$ and propositional formulas α, β,*

$$\mathrm{br}_{\mathcal{K}}(\beta) \leq \mathrm{br}_{\mathcal{K}}(\alpha).$$

Proof. Let $\mathcal{K}^* = \overline{\mathcal{E}_{\alpha}^{\mathcal{K}}}$. If $\mathrm{br}_{\mathcal{K}}(\alpha) = \infty$ then trivially the condition above is satisfied. Otherwise, suppose by contradiction that $\mathrm{br}_{\mathcal{K}}(\beta) > \mathrm{br}_{\mathcal{K}}(\alpha)$. Clearly $\mathcal{K}^* \not\models \neg\alpha$ and $\mathcal{K}^* \models \neg\beta$. Consider any model $\mathcal{I}$ of $\mathcal{K}^*$ such that $\mathcal{I}(\alpha) = \mathrm{T}$. Since $\mathcal{K}^* \models \alpha \rightarrow \beta$ by assumption, every $\mathcal{I}$ has $\mathcal{I}(\beta) = \mathrm{T}$. However, since $\mathcal{K}^* \models \neg\beta$ there can be no such $\mathcal{I}$. This means that $\mathcal{K}^* \models \neg\alpha$ and we have a contradiction.

Corollary 1. *If $\alpha \models \beta$ then for any knowledge base $\mathcal{K}$ and propositional formulas α, β,*

$$\mathrm{br}_{\mathcal{K}}(\beta) \leq \mathrm{br}_{\mathcal{K}}(\alpha).$$

Proof. This result follows from Lemma 1 above since $\alpha \models \beta$ implies that $\alpha \rightarrow \beta$ is a tautology.

Lemma 2. *For a propositional formula α and knowledge base $\mathcal{K}$, every $\mathcal{J} \in \mathcal{J}_W(\mathcal{K}, \alpha)$ has $\mathcal{J} \subseteq \mathcal{E}_{\infty}^{\mathcal{K}}$.*

Proof. In this context α is a shorthand for $\neg\alpha \mathrel{|\!\sim} \bot$. The base rank $\mathrm{br}_{\mathcal{K}}(\neg\alpha)$ of the antecedent is the smallest i with $0 \leq i \leq n$ such that $\neg\alpha$ is not exceptional for $\mathcal{E}_i^{\mathcal{K}}$ or otherwise ∞ if there is no such i. There can be no such i because if there were we would have

$$\mathrm{br}_{\mathcal{K}}(\neg\alpha) = i \text{ and } \overline{\mathcal{E}_i^{\mathcal{K}}} \not\models \neg\alpha \rightarrow \bot,$$

which would mean that $\mathcal{K} \not\approx \alpha$ (i.e. $\mathcal{K} \not\approx \neg\alpha \mathrel{|\!\sim} \bot$). Therefore $\mathrm{br}_{\mathcal{K}}(\neg\alpha) = \infty$. It follows easily that any weak justification $\mathcal{J}$ of $\mathcal{K} \approx \alpha$ has $\mathcal{J} \subseteq \mathcal{E}_{\infty}^{\mathcal{K}}$ (refer to Proposition 5).

We can now prove our results in Section 5:

Proposition 12. *If $\mathcal{D}$ is a deciding knowledge base for an entailment $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \beta$ and $\mathrm{br}_{\mathcal{K}}(\alpha) \neq \infty$, then $\mathcal{D} \approx \alpha \mathrel{|\!\sim} \beta$.*

Proof. Since $\mathrm{br}_{\mathcal{K}}(\alpha) \neq \infty$, we have $\overline{\mathcal{E}_{\alpha}^{\mathcal{K}}} \not\models \neg\alpha$. Then by classical monotonicity (or rather its contrapositive) we have $\overline{\mathcal{D}} \not\models \neg\alpha$. Hence $\mathcal{E}_{\alpha}^{\mathcal{D}} = \mathcal{D}$ and because $\mathcal{D}$ is deciding,

$$\overline{\mathcal{E}_{\alpha}^{\mathcal{D}}} \models \alpha \rightarrow \beta$$

which implies that $\mathcal{D} \approx \alpha \mathrel{|\!\sim} \beta$.

Proposition 13. *If $\mathcal{D}$ is a deciding knowledge base for an entailment $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \beta$ and we have $\mathcal{D} \approx \alpha \mathrel{|\!\sim} \beta$, then*

$$\mathcal{J}_W\left(\mathcal{D}, \alpha \mathrel{|\!\sim} \beta\right) \subseteq \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \beta\right).$$

Proof. Since $\mathcal{D}$ is deciding, we have $\mathcal{D} \subseteq \mathcal{E}^{\mathcal{K}}_{\alpha}$ and $\overline{\mathcal{D}} \models \alpha \rightarrow \beta$. All $\mathcal{J} \in \mathcal{J}_W\left(\mathcal{D}, \alpha \mathrel{|\!\sim} \beta\right)$ have $\mathcal{J} \subseteq \mathcal{D}$ and $\overline{\mathcal{J}} \models \alpha \rightarrow \beta$. Each $\mathcal{J}$ is also minimal; for every $\mathcal{J}$ there is no $\mathcal{J}' \subset \mathcal{J}$ with $\mathcal{J}' \models \alpha \rightarrow \beta$. Therefore every $\mathcal{J}$ has $\mathcal{J} \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \beta\right)$.

Theorem 2. *For any knowledge bases $\mathcal{K}, \mathcal{J}_1, \mathcal{J}_2$,*

- *(LLE)* *if $\mathcal{J}_1 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \leftrightarrow \beta\right)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \gamma\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \beta \mathrel{|\!\sim} \gamma$;*
- *(RW)* *if $\mathcal{J}_1 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \rightarrow \beta\right)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \gamma \mathrel{|\!\sim} \alpha\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \gamma \mathrel{|\!\sim} \beta$;*
- *(And)* *if $\mathcal{J}_1 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \beta\right)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \gamma\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \beta \wedge \gamma$;*
- *(Or)* *if $\mathcal{J}_1 \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \gamma\right)$ and $\mathcal{J}_2 \in \mathcal{J}_W\left(\mathcal{K}, \beta \mathrel{|\!\sim} \gamma\right)$, $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \alpha \vee \beta \mathrel{|\!\sim} \gamma$.*
- *(Ref)* *$\mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \alpha\right) = \{\emptyset\}$.*
- *(CM)* *if $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \gamma$ and $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \beta$, every $\mathcal{J} \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \gamma\right)$ is deciding for $\mathcal{K} \approx \alpha \wedge \beta \mathrel{|\!\sim} \gamma$;*
- *(RM)* *if $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \gamma$ and $\mathcal{K} \not\approx \alpha \mathrel{|\!\sim} \neg\beta$, every $\mathcal{J} \in \mathcal{J}_W\left(\mathcal{K}, \alpha \mathrel{|\!\sim} \gamma\right)$ is deciding for $\mathcal{K} \approx \alpha \wedge \beta \mathrel{|\!\sim} \gamma$.*

Proof. We consider each statement in sequence:

- (LLE) Applying Lemma 2 gives that $\mathcal{J}_1 \subseteq \mathcal{E}^{\mathcal{K}}_{\infty}$. Also notice that it follows from $\mathcal{K} \approx \alpha \leftrightarrow \beta$ that $\mathcal{K} \approx \alpha \rightarrow \beta$ and $\mathcal{K} \approx \beta \rightarrow \alpha$.[3] Applying Lemma 1 then gives $\mathrm{br}_{\mathcal{K}}(\alpha) = \mathrm{br}_{\mathcal{K}}(\beta)$. Since $\mathcal{J}_1$ and $\mathcal{J}_2$ are weak justifications and $\mathcal{J}_1 \subseteq \mathcal{E}^{\mathcal{K}}_{\infty}$, we have $\overline{\mathcal{J}_1} \models \alpha \leftrightarrow \beta$, $\overline{\mathcal{J}_2} \models \alpha \rightarrow \gamma$ and $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{E}^{\mathcal{K}}_{\alpha}$. Then $\overline{\mathcal{J}_1 \cup \mathcal{J}_2} \models \beta \rightarrow \gamma$ and $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{E}^{\mathcal{K}}_{\beta}$; hence $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \beta \mathrel{|\!\sim} \gamma$.
- (RW) Applying Lemma 2 gives that $\mathcal{J}_1 \subseteq \mathcal{E}^{\mathcal{K}}_{\infty}$. We then have $\overline{\mathcal{J}_1} \models \alpha \rightarrow \beta$, $\overline{\mathcal{J}_2} \models \gamma \rightarrow \alpha$ and $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{E}^{\mathcal{K}}_{\gamma}$. Therefore $\overline{\mathcal{J}_1 \cup \mathcal{J}_2} \models \gamma \rightarrow \beta$ and $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \gamma \mathrel{|\!\sim} \beta$.
- (And) We have $\overline{\mathcal{J}_1} \models \alpha \rightarrow \gamma$, $\overline{\mathcal{J}_2} \models \alpha \rightarrow \beta$ and $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{E}^{\mathcal{K}}_{\alpha}$. We then have $\overline{\mathcal{J}_1 \cup \mathcal{J}_2} \models \alpha \rightarrow \gamma \wedge \beta$ and therefore $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \beta \wedge \gamma$.
- (Or) Applying Corollary 1 to $\alpha \models \alpha \vee \beta$ and $\beta \models \alpha \vee \beta$ gives that $\mathrm{br}_{\mathcal{K}}(\alpha \vee \beta) \leq \mathrm{br}_{\mathcal{K}}(\alpha)$ and $\mathrm{br}_{\mathcal{K}}(\alpha \vee \beta) \leq \mathrm{br}_{\mathcal{K}}(\beta)$. We have $\overline{\mathcal{J}_1} \models \alpha \rightarrow \gamma$, $\overline{\mathcal{J}_2} \models \beta \rightarrow \gamma$ and $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{E}^{\mathcal{K}}_{\alpha} \cup \mathcal{E}^{\mathcal{K}}_{\beta}$. It follows that $\overline{\mathcal{J}_1 \cup \mathcal{J}_2} \models \alpha \vee \beta \rightarrow \gamma$ and $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{E}^{\mathcal{K}}_{\alpha\vee\beta}$; therefore $\mathcal{J}_1 \cup \mathcal{J}_2$ is deciding for $\mathcal{K} \approx \alpha \vee \beta \mathrel{|\!\sim} \gamma$.

[3] The full deduction here is $\mathcal{K} \approx \alpha \leftrightarrow \beta \implies \overline{\mathcal{E}^{\mathcal{K}}_{\infty}} \models \neg(\alpha \leftrightarrow \beta) \rightarrow \bot \implies \overline{\mathcal{E}^{\mathcal{K}}_{\infty}} \models \neg(\alpha \rightarrow \beta) \rightarrow \bot \implies \mathcal{K} \approx \alpha \rightarrow \beta$ and likewise for $\mathcal{K} \approx \beta \rightarrow \alpha$.

(Ref) We have that $\emptyset$ is a weak justification for $\mathcal{K} \approx \alpha \mathrel{|\!\sim} \alpha$ since $\emptyset \subseteq \mathcal{E}_{\alpha}^{\mathcal{K}}$ and $\emptyset \models \alpha \rightarrow \alpha$. There are no other weak justifications because any $\mathcal{K}' \neq \emptyset$ is not minimal.

(CM) Since CM is a strictly weaker condition than RM, the proof for RM applies here.

(RM) Consider any $\mathcal{J} \in \mathcal{J}_W \left(\mathcal{K}, \alpha \mathrel{|\!\sim} \gamma\right)$. We have $\overline{\mathcal{J}} \models \alpha \rightarrow \gamma$ and therefore $\overline{\mathcal{J}} \models \alpha \wedge \beta \rightarrow \gamma$. Since $\mathcal{K} \not\approx \alpha \mathrel{|\!\sim} \neg\beta$, β is not exceptional for $\mathcal{E}_{\alpha}^{\mathcal{K}}$, i.e. $\mathrm{br}_{\mathcal{K}}(\beta) \leq \mathrm{br}_{\mathcal{K}}(\alpha)$. Therefore $\mathrm{br}_{\mathcal{K}}(\alpha \wedge \beta) \geq \mathrm{br}_{\mathcal{K}}(\alpha)$ by Corollary 1. We now prove $\mathrm{br}_{\mathcal{K}}(\alpha \wedge \beta) = \mathrm{br}_{\mathcal{K}}(\alpha)$. Suppose by contradiction that $\mathrm{br}_{\mathcal{K}}(\alpha \wedge \beta) > \mathrm{br}_{\mathcal{K}}(\alpha)$. For compactness, let

$$\mathcal{K}^* = \overline{\mathcal{E}_{\alpha}^{\mathcal{K}}}.$$

Then $\mathcal{K}^* \models \neg\alpha \vee \neg\beta$ but also $\mathcal{K}^* \models \neg\alpha$ since $\mathrm{br}_{\mathcal{K}}(\alpha) \neq \infty$. This means that $\mathcal{K}^* \models \neg\beta$ but this is a contradiction since earlier we had that $\mathcal{K}^* \not\models \neg\beta$. Therefore $\mathrm{br}_{\mathcal{K}}(\alpha \wedge \beta) = \mathrm{br}_{\mathcal{K}}(\alpha)$. This implies that $\mathcal{J} \subseteq \mathcal{E}_{\alpha\wedge\beta}^{\mathcal{K}}$, and since $\overline{\mathcal{J}} \models \alpha \wedge \beta \rightarrow \gamma$, we have that $\overline{\mathcal{J}}$ is deciding for $\mathcal{K} \approx \alpha \wedge \beta \mathrel{|\!\sim} \gamma$.

References

1. S. P. Bail. *The justificatory structure of OWL ontologies.* PhD thesis, University of Manchester, 2013.
2. Or Biran and Courtenay Cotton. Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI)*, volume 8, pages 8–13, 2017.
3. Gerhard Brewka and Markus Ulbricht. Strong explanations for nonmonotonic reasoning. In *Description Logic, Theory Combination, and All That*, pages 135–146. Springer, 2019.
4. Giovanni Casini, Thomas Meyer, Kodylan Moodley, and Riku Nortjé. Relevant closure: A new form of defeasible reasoning for description logics. In Eduardo Fermé and João Leite, editors, *Logics in Artificial Intelligence*, pages 92–106, Cham, 2014. Springer International Publishing.
5. Giovanni Casini, Thomas Meyer, Kodylan Moodley, and Ivan Varzinczak. Towards practical defeasible reasoning for description logics. 2013.
6. Giovanni Casini, Thomas Meyer, and Ivan Varzinczak. Taking defeasible entailment beyond rational closure. In *European Conference on Logics in Artificial Intelligence*, pages 182–197. Springer, 2019.
7. Giovanni Casini and Umberto Straccia. Defeasible inheritance-based description logics. *Journal of Artificial Intelligence Research*, 48:415–473, 2013.
8. Victoria Chama. Explanation for defeasible entailment. Master's thesis, Faculty of Science, 2020.
9. Laura Giordano, Valentina Gliozzi, Nicola Olivetti, and Gian Luca Pozzato. Semantic characterization of rational closure: From propositional logic to description logics. *Artificial Intelligence*, 226:1–33, 2015.
10. Matthew Horridge. *Justification based explanation in ontologies.* The University of Manchester (United Kingdom), 2011.

11. Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1):167–207, 1990.
12. Daniel Lehmann. Another perspective on default reasoning. *Annals of mathematics and artificial intelligence*, 15(1):61–82, 1995.
13. Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial intelligence*, 55(1):1–60, 1992.
14. Kody Moodley. *Practical reasoning for defeasable description logics.* PhD thesis, University of KwaZulu-Natal, 2016.
15. Matthew Morris, Tala Ross, and Thomas Meyer. Algorithmic definitions for klm-style defeasible disjunctive datalog. *South African Computer Journal*, 32(2):141–160, 2020.