An Overview of KLM-Style Defeasible Entailment

By

Adam Kaliski

A thesis presented in accordance with the requirements for the degree of Masters of Science in the Faculty of Science

Declaration

I, Adam Kaliski, do hereby declare that this Masters thesis titled "An Overview of KLM-Style Defeasible Entailment" is my own work and has not been submitted for any other degree award to any other University. I understand that failure to attribute material which is obtained from another source may be considered plagiarism.

Signed: Signed by candidate

Date: 16/11/2020

Acknowledgements

First and foremost, thank you to my supervisor, Thomas Meyer, for his guidance and mentorship. At every stage of this project, his support was crucial and comprehensive. Without his input, my academic progress would be less than half what it is now.

I would also like to thank Giovanni Casini and Ivan Varzinczak for their valuable tutelage and input over the course of this project.

Thank you to the Centre for Artificial Intelligence Research (CAIR), for their financial support, without which this project would have been far harder to complete.

Thank you to the staff and fellow students at the UCT computer science department for their community and feedback.

Finally, thank you to my friends and family for their support, in various ways, over the course of this project.

Contents

Abstract

The usage of formal logic to solve problems in artificial intelligence has a long history in the field. Information is represented in a formal language, which facilitates algorithmic reasoning about some domain knowledge. Traditionally, the algorithms used for the reasoning services are *monotonic*, which states that adding knowledge never causes the retraction of an inference. A result of this is that if the knowledge in question contains examples that are exceptions to stated rules, then the entire knowledge base may become unsatisfiable. If the knowledge accurately represents the domain, then such a result is undesirable. One solution is *nonmonotonic* reasoning, which encompasses patterns of defeasible or "common sense" reasoning that may retract conclusions upon the addition of new information to a knowledge base. One of the most prominent frameworks for nonmonotonic reasoning is the one defined by Kraus, Lehmann, and Magidor (KLM). The KLM framework has very desirable features both for theoretical study of nonmonotonic reasoning, as well as for implementation in AI applications. However, the current state of the KLM framework spans numerous papers over two decades of research. This provides a challenge for new researchers to understand the current problems being studied, as well as to understand the framework well enough to either extend it or apply it. This dissertation aims to compile the theoretical work done in this framework to provide a single point of reference for anyone wishing to understand the KLM framework, as well as to know how to define a defeasible entailment relation, using homogenised terminology and notation that is now typical of the field. Firstly, the propositional logic used as the base language will be defined. Then, paralleling the way the framework was historically built up, a preferential semantics over that language will be described, before modifying the language itself with a defeasible connective, and introducing a nonmonotonic entailment relation over such a language. Then, recent extensions to this framework defining various classes of defeasible entailment are described. By the end of this dissertation, the reader should have a well rounded understanding of the KLM framework, from classical logic to defeasible logic.

Chapter 1

Introduction and Background

Knowledge representation and reasoning is a field of artificial intelligence that employs formal logic to represent domain knowledge symbolically, which allows for rigorous analysis of additional knowledge, inferences, implied from the explicit knowledge [5, 35]. This analysis can be defined algorithmically, and therefore implemented in the form of reasoning services. These reasoning services are intended to check for consistency, that there are no conflicts in the information provided, as well as provide inferences. The set of inferences derived by a particular reasoning algorithm is dependent on the assumptions baked into the algorithm. The dominant form of reasoning, referred to as *classical* reasoning, is based on Tarskian notions of consequence [56], which define key properties to be satisfied such that each inference has maximal support from the explicit knowledge. A side-effect is that classical notions of consequence cannot reason about exceptional explicit knowledge. In this context, *exceptional* refers to information in which there may be examples of exceptions to stated rules. To use the typical example, a bird has wings and flies, and a penguin is a bird that does not fly. Therefore, a penguin is an *exceptional* bird. In propositional logic, this knowledge could be formalised as:

- $\texttt{bird} \rightarrow \texttt{flies}$
- $\texttt{bird} \rightarrow \texttt{wings}$
- $\texttt{penguin} \rightarrow \texttt{bird}$
- $\texttt{penguin} \rightarrow \neg\texttt{flies}$

Using the regular, classical rules of drawing inferences, this information infers both that `penguin` $\rightarrow$ $\neg$`flies`, explicitly, and that `penguin` $\rightarrow$`flies`, derived from the fact that it is a `bird`. This is an undesirable result, as it implies that there are no penguins, as either there exists a `bird` with contradictory traits, resulting in an explicit contradiction, or there exists no such `bird`. It is possible to adjust the above knowledge base to account for penguins as an exception. However, adding ostriches to the knowledge base will require the same handling, and so the more exceptions there are, the more adjustments are required, and the larger the knowledge base becomes and the harder it is to maintain. Given an existing knowledge base with many hundreds or thousands of statements, and it may not be at all reasonable to adjust the explicit knowledge in such a way as to handle such exceptions.

This issue of exceptional information is a direct result of classical reasoning satisfying the property of *monotonicity*. Monotonicity essentially states that the inferences that hold within any subset of a set of statements, have to be consequences of the entire set. Another way of stating monotonicity is that any expansion of some knowledge can only add more inferences, and can never cause a retraction of any inference. Monotonicity is therefore why in the above example the inference `penguin` $\rightarrow$ `flies` must always be derived in the presence of the statements `bird` $\rightarrow$ `flies` and `penguin` $\rightarrow$ `bird`, regardless of any other statements. The solution proposed here is therefore that of *nonmonotonic* reasoning.

Nonmonotonic reasoning is an area of research that attempts to formalize different patterns of "common sense" reasoning, by dropping monotonicity as a property and investigating how to define reasonable non-Tarskian notions of consequence. Generally, humans reason by making assumptions based on their given knowledge, and then revising those assumptions upon learning new information. This pattern of *defeasible* reasoning has been mimicked by a number of different frameworks that were formalized, for the most part, in the 1980s and 1990s. The aim of this dissertation is to provide an overview of one of these frameworks, first defined by Kraus, Lehmann and Magidor [46, 48, 49]. Propositional logic will be the basis used for this dissertation, in keeping with the language used in the original definitions. However, this framework has been extended to both description logics and modal logics [19, 20, 21, 22, 25, 37, 38].

1.1 Dissertation Outline

Chapter 2 of this dissertation defines the propositional logic that will form the foundational logic for the rest of this dissertation. The semantics and proof theory are defined, as is the concept of consequence relations. Chapter 3 is a literature review for nonmonotonic logic, covering a number of various frameworks. Chapter 4 then describes preferential logic, starting with defining preferential consequence relations and the associated preferential semantics. Then, preferential entailment over a defeasible logic is defined, with the corresponding semantics. Chapter 5 then introduces nonmonotonic rational entailment relations, starting from rational closure, and then builds up the framework by defining iterative classes of defeasible entailment. The final chapter contains a discussion on nonmonotonicity, before concluding with future work in this area.

Chapter 2

Propositional Logic

Propositional logic is a formalism for reasoning about knowledge or information, abstracted away from a natural language representation into a formal language, and will form the foundational logic for the rest of this dissertation. Propositional logic is essentially defined by a set of connectives along with a set of statements, such that more complicated statements may be constructed by combining statements using the connectives; the truth of the composite statement is then solely reliant on the truth of the base statements and the interpretation of the connectives used [5, 35]. In this way, propositional logic attempts to analyse the truth of statements divorced from their natural language intuitions. This also allows for defining what it means to formally reason about some set of such statements, and what can be derived from a set of such statements.

Propositional logic was chosen as the base logic as it was the foundational logic chosen by Kraus, Lehmann and Magidor to initially define the KLM framework [46, 49]. It is an easy choice to motivate in general in the context of artificial intelligence research, as it is decidable but also expressive enough for results to be translated to more expressive languages. First, the basic language and syntax will be defined, then the model and proof theory.

2.1 Syntax

The language of propositional logic is built up from propositional *atoms*, also called variables or symbols in the literature [35, 63]. A propositional atom is a statement, or a representation of a statement, and will be formatted here in typewriter text, e.g. `bird` is an atom, as is `Socrates`. The intuition is that `Bird` could be shorthand for a statement, or proposition, such as "Tweety is a bird", and, similarly, `Socrates` might be shorthand for "there is a philosopher named Socrates". Atoms are associated with truth values in the sense that they can be true, referred to as T, or false, referred to as F. Statements will also be represented using meta variables, denoted with small Latin alphabet letters: p, q, r... such that, for example, `bird` could be denoted as b, without any loss of interpretation. The set of all propositional atoms will be denoted as P, and is finite. Atoms are combined with the set of connectives $\{\neg, \wedge, \vee, \rightarrow, \leftrightarrow\}$ to create a set of *well formed formulas*, from now referred to as propositional formulas, or just formulas. All of the above connectives are binary, with the exception of $\neg$, which is unary, and each accepts all formulas as arguments. The set of all formulas will be denoted by $\mathcal{L}$, and elements of $\mathcal{L}$ will be denoted by lower case Greek letters e.g. α, β, γ.... The definition of a formula is recursive, such that any $\mathcal{L}$ can be defined as follows: for some $p \in P$ and $\alpha, \beta \in \mathcal{L}$, then $\alpha : p$, $\neg\alpha$, $\alpha \wedge \beta$, $\alpha \vee \beta$, $\alpha \rightarrow \beta$, $\alpha \leftrightarrow \beta$. Lastly, define the constants $\top \in \mathcal{L}$ and $\bot \in \mathcal{L}$: $\top$, read "top", is a tautology, i.e., a statement that is always true, and $\bot$ is the opposite, that is a statement that is always false. $\mathcal{L}$ then defines the syntax of propositional logic, but how to determine the truth of a formula is not yet clear, nor the ultimate goal: what new formulas may be inferred from some set of formulas. There are a number of ways to define inference in propositional logic: here a model-theoretic definition, using semantics, is presented first, followed by different, associated formalisms.

2.2 Semantics

The semantics of any logic defines the meaning of truth and allows for systematic, meaningful analysis of the language. To use the example of ordinary arithmetic, a sentence $x+y = z$ is neither true nor false until you assign values to x, y and z. In this case, it is true in the case where $\{x = 3, y = 3, z = 6\}$ and false in the case where $\{x = 4, y = 5, z = 7\}$ [5]. Each of the above assignments to x, y and z are mappings from x, y and z to the set of integers, and provide a lens to systematically analyse under what conditions the statement holds. Using the above two assignments, for example, shows that it is possible for the statement to be true or false. Satisfaction is the notion of a statement being true in such an assignment. In this example, the first assignment satisfies the statement, and the second does not. It should be noted that these assignments map to integers, not truth values, which will be contrasted with the assignments for propositional logic. Valuations, also referred to as *worlds* or *interpretations* in the literature [5, 46], assign truth to propositional atoms [51], and will be denoted by the small Latin alphabet letters u, v, w:

Definition 2.1. *A valuation u is a function, such that $u : P \mapsto \{T, F\}$ where T and F are read as true and false, respectively.*

That is, a valuation is a function that assigns for every atom in the language either *true*, or *false*. Here, a valuation will be represented as a sequence of atoms in typewriter text, where a bar over the atom is taken to mean that said atom is false in the valuation, and true otherwise. For example, given $P \coloneqq \{p, q, r\}$, a random valuation $u \in \mathcal{U}$ could be $p\bar{q}r$, which should be read as p and r are true, and q false.

Definition 2.2. *If an atom is true in a valuation, then it is said the valuation satisfies the atom, and satisfaction will be denoted with $\Vdash$, such that for some valuation u, if it is the case that for some $p \in P$, $u(p) = T$, then $u \Vdash p$, and if $u(p) = F$, then $u \nVdash p$. This can be extended to any $\alpha, \beta \in \mathcal{L}$, such that a valuation u can be said to satisfy any formula in $\mathcal{L}$ using the following criteria:*

- $u \Vdash \neg\alpha$ *if and only if* $u \nVdash \alpha$
- $u \Vdash \alpha \wedge \beta$ *if and only if* $u \Vdash \alpha$ *and also* $u \Vdash \beta$
- $u \Vdash \alpha \vee \beta$ *if and only if either* $u \Vdash \alpha$ *or* $u \Vdash \beta$
- $u \Vdash \alpha \rightarrow \beta$ *if and only if* $u \nVdash \alpha$ *or* $u \Vdash \beta$
- $u \Vdash \alpha \leftrightarrow \beta$ *if and only if* $u(\alpha) = u(\beta)$
- $u \Vdash \top$ *for every* $u \in \mathcal{U}$
- $u \nVdash \bot$ *for every* $u \in \mathcal{U}$

Now, define $\mathcal{U}$ to be the set of all valuations for our language $\mathcal{L}$. It will be useful to refer to only those valuations satisfying a particular formula, hence the following definition:

Definition 2.3. *For any* $\alpha \in \mathcal{L}$*, let* $\hat{\alpha} \coloneqq \{u \in \mathcal{U} \mid u \Vdash \alpha\}$*. For any* $u \in \hat{\alpha}$*,* u *is referred to as a model of* α*.*

Then, $\hat{\alpha}$ is a subset of $\mathcal{U}$, referring to only those valuations satisfying α. For any $\alpha \in \mathcal{L}$, α is called a *tautology* if for all $u \in \mathcal{U}$, it is the case that $u \Vdash \alpha$, α is said to be satisfiable if *there exists* a $u \in \mathcal{U}$ such that $u \Vdash \alpha$. A formula α is consequently said to be unsatisfiable if there does not exist a $u \in \mathcal{U}$ such that $u \Vdash \alpha$.

The notion of satisfaction can be used to define *logical consequence*:

Definition 2.4. *Given two formulas* $\alpha, \beta \in \mathcal{L}$*, the relation* $\alpha \vDash \beta$*, read that* β *is a logical consequence of* α*, holds if and only if for every* $u \in \mathcal{U}$ *such that* $u \Vdash \alpha$*, then* $u \Vdash \beta$*, or, equivalently,* $\alpha \vDash \beta$ *if and only if* $\hat{\alpha} \subseteq \hat{\beta}$*.*

Logical consequence then leads to the following definition of logical equivalence:

Definition 2.5. *Given two formulas* $\alpha, \beta \in \mathcal{L}$*, the relation* $\alpha \equiv \beta$ *holds if and only if* $\alpha \vDash \beta$ *and also* $\beta \vDash \alpha$*.*

Logical consequence, and logical equivalence, are both *meta-level* concepts: they are not defined as a part of $\mathcal{L}$ itself, but rather represent deductions that can be made about statements in the language.

The above semantic concepts can also be extended to sets of statements:

Definition 2.6. *Let* $\mathcal{K}$ *be a finite set of propositional formulas. A given valuation,* $u \in \mathcal{U}$*, then satisfies* $\mathcal{K}$*,* $u \Vdash \mathcal{K}$ *if and only if it is the case that for every formula* $\alpha \in \mathcal{K}$*, it is the case that* $u \Vdash \alpha$*.*

Much like with a single formula, it is then said that u is a model of $\mathcal{K}$. Then, classical entailment can be defined as logical consequence from a set of statements:

Definition 2.7. *Given* $\mathcal{K}$ *a set of formulas, and a formula* α*, it is the case that* α *is entailed by* $\mathcal{K}$*, written* $\mathcal{K} \vDash \alpha$*, if and only if for every* $u \in \mathcal{U}$ *such that* $u \Vdash \mathcal{K}$*, it is the case that* $u \Vdash \alpha$*.*

In natural language, $\mathcal{K}$ entails any α if every model of $\mathcal{K}$ is also a model of α. Entailment is essentially a pattern of reasoning that defines what formulas follow from a set of formulas. The semantics defines this model of reasoning in an unambiguous way: the above definition states that some formula follows from a set of formulas if it is true in every case where the knowledge base is true. Note that another way of defining entailment is that for any set of propositional formulas $\mathcal{K}$ and a given formula α, $\mathcal{K} \vDash \alpha$ if and only if $\hat{\mathcal{K}} \subseteq \hat{\alpha}$, that is if the set of models of $\mathcal{K}$ is contained in the models of α. This inherently ties entailment to finding models of a formula, said to be the satisfiability problem, which is NP-complete, however there are algorithms that can efficiently produce an answer for many cases [5].

For an example, let $P = \{p, q\}$. Then the set of all valuations will be $\mathcal{U} \coloneqq \{\mathsf{pq}, \mathsf{p\overline{q}}, \mathsf{\overline{p}q}, \overline{\mathsf{pq}}\}$. Recall that in each valuation p asserts that p is true, and $\overline{\mathsf{p}}$ asserts that p is false. Then, for example, in the valuation $\mathsf{\overline{p}q}$, p is false, and q is true. Now, given a $\mathcal{K} = \{p \rightarrow q\}$, it should be noted that not all valuations satisfy $\mathcal{K}$. The valuation $\mathsf{p\overline{q}}$ does not satisfy $\mathcal{K}$ because while p is true, q is false, which contradicts our propositional statement $p \rightarrow q$, which requires q to be true whenever p is true. This leaves the remaining valuations: $\mathsf{pq}, \mathsf{\overline{p}q}, \overline{\mathsf{pq}}$ as the set of valuations satisfying $\mathcal{K}$. Therefore, the set of valuations $\mathsf{pq}, \mathsf{\overline{p}q}, \overline{\mathsf{pq}}\}$ are referred to as the models of $\mathcal{K}$.

Logical consequence of a set of formulas also informs logical equivalence of sets of formulas:

Definition 2.8. *Given $\mathcal{K}_1, \mathcal{K}_2 \subseteq \mathcal{L}$, the relation $\mathcal{K}_1 \equiv \mathcal{K}_2$ holds if and only if for every $u \in \mathcal{U}$ such that $u \Vdash \mathcal{K}_1$, then $u \Vdash \mathcal{K}_2$ and for every $v \in \mathcal{U}$ such that $v \Vdash \mathcal{K}_2$ then $v \Vdash \mathcal{K}_1$.*

The above definition simply states that logically equivalent sets of formulas share the exact same models. It naturally follows that logically equivalent sets of formulas entail the exact same formulas.

Entailment is a meta-level notion, exactly like logical consequence and equivalence. The notion of $\vDash$ is not defined in $\mathcal{L}$, instead it is the concept of what can be inferred from formulas in $\mathcal{L}$. A significant implication of this is that while the above definition of entailment is a classical, Tarskian definition of consequence [56, 67], it is not the only definition for entailment, as will be elaborated on in later chapters. The above notion of entailment can be used to define a *consequence operator*:

Definition 2.9. *A consequence operator for $\mathcal{L}$, $\mathcal{C}n$ is a function such that: $\mathcal{C}n : 2^{\mathcal{L}} \mapsto 2^{\mathcal{L}}$.*

Consequence operators receive as input a set of formulas, and output a set of formulas. The intuition is that a consequence operator represents some notion of consequence: from a set of formulas as input, another set of output formulas is inferred. Consequence operators have a flexible definition by design: they can be used to represent any type of reasoning. Using the definition of entailment above, a corresponding consequence operator, $\mathcal{C}n_{\models}$ can be defined as follows: $\mathcal{C}n_{\models}(\mathcal{K}) = \{\alpha \mid \mathcal{K} \models \alpha\}$, and likewise the reverse can be formalized: $\mathcal{K} \models \alpha$ if and only if $\alpha \in \mathcal{C}n_{\models}(\mathcal{K})$.

Generally, when discussing sets of formulas, the term used is *knowledge base*:

Definition 2.10. *A knowledge base, $\mathcal{K} \subseteq \mathcal{L}$, is a finite set of propositional formulas.*

A knowledge base is intended to represent a set of facts about the world, or some domain. For example, the knowledge base $\mathcal{K} = \{p, p \to q\}$ states that there are two known facts: that p and $p \to q$ both hold. What exactly p and q are is unimportant. Rather, each formula in a knowledge base can be thought of as restricting the set of valuations in $\mathcal{U}$ that satisfy the knowledge base. Above, the notions of entailment and satisfaction that applied to some set of formulas $\mathcal{K}$, naturally all apply in general to any knowledge base.

It is possible that for every $u \in \mathcal{U}$ that $u \not\Vdash \mathcal{K}$, or that there is no model of $\mathcal{K}$. In this case, the knowledge base is unsatisfiable, and generally implies that a contradiction that is logically equivalent to $\alpha \wedge \neg\alpha$ is derivable from the knowledge base. If this is the case, then the set of valuations satisfying $\mathcal{K}$ is the empty set, $\varnothing$, and therefore for every formula $\alpha \in \mathcal{L}$, it is the case that $\hat{\mathcal{K}} \subseteq \hat{\alpha}$. Therefore, any unsatisfiable knowledge base entails the entire language. This result is known more generally as the axiom of explosion, and can be formalized as follows:

Corollary 2.0.1. *Given a knowledge base $\mathcal{K}$, if for all $u \in \mathcal{U}$ it is the case that $u \not\Vdash \mathcal{K}$ then it is the case that $\mathcal{K} \models \alpha$ for any $\alpha \in \mathcal{L}$.*

Defining the semantics with respect to a knowledge base allows for satisfiability checking, but reasoning is broader than that. A *query* is any propositional formula, and query checking is the process of determining if a particular formula is entailed by a knowledge base. This is no more difficult a task to satisfiability checking, as query checking is analogous to checking if the knowledge base is unsatisfiable when the negation of a formula is added to the knowledge base.

Lemma 2.0.1. *Given a knowledge base $\mathcal{K}$, and a query $\alpha \in \mathcal{L}$, the relation $\mathcal{K} \models \alpha$ holds if and only if $\mathcal{K} \cup \{\neg\alpha\} \models \bot$, i.e., if $\mathcal{K} \cup \{\neg\alpha\}$ is unsatisfiable.*

With that in mind, extending the previous example, let the set of propositional atoms of our logic be $P \coloneqq \{p, q, r\}$ and let the knowledge base $\mathcal{K} \coloneqq \{p \rightarrow q\}$. Now, consider the following query: $p \wedge r \rightarrow q$. The set of all valuations of P is the following set $\mathcal{U} = \{\mathtt{pqr}, \mathtt{p\overline{q}r}, \mathtt{\overline{p}qr}, \mathtt{\overline{pq}r}, \mathtt{pq\overline{r}}, \mathtt{\overline{p}q\overline{r}}, \mathtt{p\overline{qr}}, \mathtt{\overline{pqr}}\}$. Then, the set of valuations satisfying $\mathcal{K}$ is $\hat{\mathcal{K}} = \{\mathtt{pqr}, \mathtt{\overline{p}qr}, \mathtt{\overline{pq}r}, \mathtt{pq\overline{r}}, \mathtt{\overline{p}q\overline{r}}, \mathtt{\overline{pqr}}\}$, and the set of valuations satisfying the query statement $p \wedge r \rightarrow q$ is $\{\mathtt{pqr}$, $\mathtt{\overline{p}qr}$, $\mathtt{\overline{pq}r}$, $\mathtt{pq\overline{r}}$, $\mathtt{\overline{p}q\overline{r}}$, $\mathtt{p\overline{qr}}$, $\mathtt{\overline{pqr}}\}$. If $\hat{\mathcal{K}} \subseteq \hat{\alpha}$ then $\mathcal{K} \vDash \alpha$, and in this case $\hat{\mathcal{K}} \subseteq \hat{\alpha}$ is true, and therefore it is true that $\mathcal{K} \vDash p \wedge r \rightarrow q$.

2.3 Deductive Systems

As a complement to the semantic definition of entailment, another methodology to draw inferences from a knowledge base is that of *deductive systems* [5]. Generally speaking, deductive systems are defined with respect to a semantics. Whereas semantics declaratively define the notion of truth in a logic, and therefore logical consequence, deductive systems allow for a mathematical, or algorithmic, definition of logical consequence. The core idea is to define what it means for an inference to follow from explicit knowledge by defining rules that govern the pattern of reasoning. There are a number of practical reasons to define deductive systems: if the explicit information is very large then it is computationally expensive, if possible, to compute the inferences purely semantically, whereas a deductive system would allow for inferences to be derived relatively cheaply. Another motivation is that intermediate results, lemmas, are lost in a purely semantic system, as the only output is either *true* or *false*, while a deductive system shows each step of reasoning from explicit to implicit information.

A core question of defining such a syntactic formulation of reasoning is, naturally, how to define logical consequence syntactically. The answer is the formalism of *rules of inference* [69]:

Definition 2.11. *A rule of inference, $r \coloneqq \{Q, c\}$, is a set of formulas, $Q = \{q_1, ...q_n\}$, called the premises and a formula, c, called the conclusion. A given rule of inference, r, will have the following syntax: $r = \frac{q_1, \ldots, q_n}{c}$.*

Rules of inference are syllogistic in nature, and represent a basic pattern of deduction: given that the premises are true, then the conclusion is derivable. A rule of inference is sound with respect to a logic, if it corresponds to the existing notions of logical consequence.

Definition 2.12. *A rule of inference* $\frac{q_1,\ldots,q_n}{c}$ *is sound if and only if* $q_1, ..., q_n \vDash c$.

The other component of deductive systems are axiom schemas:

Definition 2.13. *An axiom schema is a meta formula comprised of meta variables, representing all formulas in the language of the same structure.*

The exact structure of an axiom schema is dependent on the deductive system itself. Axiom schemas may be purely an object-level statement, or a meta-level one.

Having defined rules of inference and axiom schemas, deductive systems can then be defined [5, 69]:

Definition 2.14. *A deductive system for propositional logic is* $\mathcal{D} \coloneqq \langle LA, \mathcal{R} \rangle$, *with* LA *a set of axiom schemas, and* $\mathcal{R}$ *a set of rules of inference.*

Deductive systems either tend to have many rules and few axioms, or vice versa. For example, Gentzen systems have one type of axiom schema, and many rules of inference, whereas Hilbert systems have one rule of inference and several axiom schema [5].

In addition to axiom schema, there are what are referred to as specific axioms [69]. Specific axioms are the formulas in a knowledge base $\mathcal{K}$. The knowledge base $\mathcal{K}$, along with the axiom schema, and rules of inference are used to define proofs:

Definition 2.15. *Given a deductive system* $\mathcal{D}$, *and a set of specific axioms* $\mathcal{K}$, *a proof in* $\mathcal{D}$ *is a sequence of formulas* $S = \{\alpha_1, ..., \alpha_n\}$ *where every formula* α_i *is either an instance of an axiom schema or a formula* $\alpha \in \mathcal{K}$, *or is derived from some subset of the previous formulas* $\alpha_1, ..., \alpha_j$, *where* $1 \leq j < i$, *using a rule of inference. The last formula,* α_n, *of a proof is referred to as provable in* $\mathcal{D}$ *from* $\mathcal{K}$, *denoted* $\mathcal{K} \vdash_{\mathcal{D}} \alpha_n$.

The last formula of a proof may be used as an axiom in some subsequent proof from $\mathcal{K}$ using the same deductive system.

A natural question is how to verify whether a deductive system accurately models the kind of reasoning defined by the semantics in the previous section. Recall that a rule of inference is sound with respect to a semantics if it corresponds to the notion of logical consequence in such a semantics. This can be extended to the whole deductive system. That is, a deductive system,

$\mathcal{D}$, is sound with respect to a semantics if every formula true in $\mathcal{D}$ is also true in the semantics, and complete if every formula true in the semantics is true in $\mathcal{D}$. If soundness and completeness are proven for $\mathcal{D}$, then $\mathcal{D}$ exactly represents the reasoning pattern defined by the semantics of a given logic.

2.3.1 Hilbert System

The first deductive system chosen to be described here is the Hilbert System, $\mathcal{H}$ [5]. Hilbert systems exist for many logics, naturally here it will be described with reference to propositional logic.

First, some definitions to clarify axiom schemas in the context of this Hilbert system:

Definition 2.16. *An axiom schema for system $\mathcal{H}$ is a meta formula α such that α may be replaced by any formula $\beta \in \mathcal{L}$ of the same form.*

The notion of β being of the same form as α is essentially that β has the same structure as α. For an example, an axiom schema could be $\beta \to \alpha$, such that β and α are then meta formulas, and can be replaced by any $\gamma \to \delta \in \mathcal{L}$.

As mentioned, Hilbert systems have many axiom schemas and a single rule of inference. For propositional logic, they are as follows:

Definition 2.17. *The Hilbert system $\mathcal{H}$ logical axiom schemas are:*

1. $\alpha \to (\beta \to \alpha)$
2. $(\alpha \to (\beta \to \gamma)) \to ((\alpha \to \beta) \to (\alpha \to \gamma))$
3. $(\neg\beta \to \neg\alpha) \to (\alpha \to \beta)$

along with the following rule of inference, modus ponens (MP): $\dfrac{\alpha, \alpha \to \beta}{\beta}$

For any knowledge base $\mathcal{K}$, if a formula $\alpha \in \mathcal{L}$ has a proof from $\mathcal{K}$ in $\mathcal{H}$ then it is denoted $\mathcal{K} \vdash_{\mathcal{H}} \alpha$.

The Hilbert system only has one rule of inference in its definition, *modus ponens*, but there are a number of rules of inference that may be derived. An example is the deduction rule [5]:

Theorem 2.1. *The deduction rule,* $\dfrac{\mathcal{K} \cup \{\alpha\} \vdash \beta}{\mathcal{K} \vdash \alpha \to \beta}$*, is a sound derived rule in system $\mathcal{H}$ for propositional logic.*

There are many such derivable rules in $\mathcal{H}$, built upon previously derived rules. However, the axiom schemas and *modus ponens* is all that is needed for the definition.

Hilbert systems for propositional logic are sound and complete with respect to the semantics for propositional logic:

Theorem 2.2 (Soundness and completeness)**.** *Given a knowledge base $\mathcal{K}$ and any $\alpha \in \mathcal{L}$, the relation $\mathcal{K} \vdash_{\mathcal{H}} \alpha$ holds if and only if $\mathcal{K} \vDash \alpha$.*

The proofs for soundness and completeness for a Hilbert system for propositional logic are well established in the literature [5, 69].

To illustrate the pattern of a formal proof, the formula $\neg\alpha \rightarrow (\alpha \rightarrow \beta)$ will be shown to be a tautology, i.e., derivable from the given logical axioms and rules with no specific axioms:

1. $\vdash \neg\alpha \rightarrow (\neg\beta \rightarrow \neg\alpha)$ — Axiom 1
2. $\neg\alpha \vdash \neg\beta \rightarrow \neg\alpha$ — Deduction rule
3. $\vdash (\neg\beta \rightarrow \neg\alpha) \rightarrow (\alpha \rightarrow \beta)$ — Axiom 3
4. $\neg\alpha \vdash \alpha \rightarrow \beta$ — MP 2,3
5. $\vdash \neg\alpha \rightarrow (\alpha \rightarrow \beta)$ — Deduction rule

Proofs using any deductive system follow the above pattern. Note that every line is either an axiom, specific or logical, or an application of a rule of inference to a subset of previous lines.

2.3.2 Gentzen System

The other deductive system presented here is a Gentzen system for propositional logic. As before, first the axiom schemas in this system are defined as follows:

Definition 2.18. *An axiom schema for the Gentzen system is a statement $\alpha \vdash \beta$ where α and β are meta-formulas that may be replaced by any $\gamma, \delta \in \mathcal{L}$ of the same form as α and β.*

A meta formula α may be replaced by a formula $\gamma \in \mathcal{L}$ of the same form in the same way as before: provided the structure of γ is identical to the structure of α.

Similarly, the definition given above for rules of inference needs to be refined for Gentzen systems:

Definition 2.19. *Given $\Gamma, \Delta, \Phi, \Psi \subseteq \mathcal{L}$ as sets of formulas, then a rule of inference for the Gentzen system takes the form:* $\frac{\Gamma \vdash \Delta}{\Phi \vdash \Psi}$.

The difference to note is that the rules of inference, as well as the axiom schema, in Gentzen systems use meta statements: statements containing the meta symbol $\vdash$, used to denote logical consequence, rather than being restricted purely to the object-level. The basic concept is identical however: given that it is the case that $\Gamma \vdash \Delta$ then $\Phi \vdash \Psi$ is derivable.

In contrast to Hilbert systems, Gentzen systems have many rules of inference, and one axiom schema. The following is a formal definition of a Gentzen system for propositional logic [4].

Definition 2.20. *A Gentzen proof system for propositional logic consists of the axiom schema:*

$$\alpha \vdash \alpha$$

and the following rules of inference, where $\mathcal{K}, \Delta \subseteq \mathcal{L}$ are sets of formulas, and $\alpha, \beta \in \mathcal{L}$ are single formulas:

- $\frac{\mathcal{K} \vdash \Delta \cup \{\alpha\}}{\{\neg\alpha\} \cup \mathcal{K} \vdash \Delta}$ *Left negation*
- $\frac{\{\alpha\} \cup \mathcal{K} \vdash \Delta}{\mathcal{K} \vdash \Delta \cup \{\neg\alpha\}}$ *Right negation*
- $\frac{\mathcal{K} \vdash \Delta \cup \{\alpha\} \quad \{\beta\} \cup \mathcal{K} \vdash \Delta}{\{\alpha \rightarrow \beta\} \cup \mathcal{K} \vdash \Delta}$ *Left implication*
- $\frac{\mathcal{K} \cup \{\alpha\} \vdash \Delta \cup \{\beta\}}{\mathcal{K} \vdash \Delta \cup \{\alpha \rightarrow \beta\}}$ *Right implication*
- $\frac{\mathcal{K} \cup \{\alpha\} \cup \{\beta\} \vdash \Delta}{\mathcal{K} \cup \{\alpha \wedge \beta\} \vdash \Delta}$ *Left conjunction*
- $\frac{\mathcal{K} \vdash \Delta \cup \{\alpha\} \quad \mathcal{K} \vdash \Delta \cup \{\beta\}}{\mathcal{K} \vdash \Delta \cup \{\alpha \wedge \beta\}}$ *Right conjunction*

- $\frac{\mathcal{K}\cup\{\alpha\}\vdash\Delta \quad \mathcal{K}\cup\{\beta\}\vdash\Delta}{\mathcal{K}\cup\{\alpha\vee\beta\}\vdash\Delta}$ *Left disjunction*

- $\frac{\mathcal{K}\vdash\Delta\cup\{\alpha\}\cup\{\beta\}}{\mathcal{K}\vdash\Delta\cup\{\alpha\vee\beta\}}$ *Right disjunction*

For any knowledge base $\mathcal{K}$, if α has a proof from $\mathcal{K}$ then it is denoted $\mathcal{K} \vdash_{\mathcal{G}} \alpha$.

The above Gentzen system is sound and complete with respect to the semantics of propositional logic [35]:

Theorem 2.3 (Soundness and completeness)**.** *Given a knowledge base $\mathcal{K}$ and a formula $\alpha \in \mathcal{L}$, then $\mathcal{K} \vdash_{\mathcal{G}} \alpha$ if and only if $\mathcal{K} \vDash \alpha$.*

For an example of a proof using a Gentzen system, consider the following proof of the tautology $\vdash (p \wedge q) \rightarrow (q \wedge p)$. Note that for any $\alpha, \beta \in \mathcal{L}$, then α, β is a syntactic shorthand for $\{\alpha\} \cup \{\beta\}$:

1. $p, q \vdash p, q$ Axiom
2. $\vdash \neg p, \neg q, q, p$ Right negation
3. $\vdash \neg(p \wedge q), q, p$ Right disjunction
4. $\vdash \neg(p \wedge q), (q \wedge p)$ Right disjunction
5. $p \wedge q \vdash q \wedge p$ Left negation
6. $\vdash (p \wedge q) \rightarrow (q \wedge p)$ Right implication

2.4 Consequence Relations

A more abstract notion of defining logical consequence, consequence relations are structures defining a pattern of reasoning over a given language. Rather than working directly on a given knowledge base, as in semantics, or in deductive systems where they were referred to as specific axioms, consequence relations define logical consequence as a mathematical relation between formulas, such that the relation conforms to some set of properties constraining what pattern of reasoning is represented by the relation.

Formally, a consequence relation is a, possibly infinite, set of ordered pairs: $\{(\alpha_1, \beta_1), ..., (\alpha_n, \beta_n), ...\}$ with each $\alpha_i, \beta_i \in \mathcal{L}$. Generally, it is accepted that each ordered pair may contain a set of formulas as the antecedent, that is α_i, and a single formula as the consequent, β_i. Here, the consequence relations that will be considered are those where the antecedent, α_i, is a single formula. Consequence relations are often denoted as a binary relation, $\vdash$, such that $\alpha \vdash \beta$ is interpreted that (α, β) is in the set and $\alpha \nvdash \beta$ is interpreted that it is not. The intuition meant to be represented by a consequence relation is that each ordered pair represents an inference. If (α, β) is in the consequence relation, the meaning attached is that β can be inferred from α, and this meaning should then also be attached to the binary relation, with $\alpha \vdash \beta$ meaning that β can be inferred from α.

It should be clear that not all such sets of ordered pairs represent a meaningful notion of inference. It is simple to construct a random such set, e.g. $\{(\alpha, \beta), (\gamma, \delta)\}$ which represents no meaningful notion of consequence in isolation, regardless of the interpretation of the formulas. Avron [3] described two main properties that a consequence relation should satisfy to represent a reasonable notion of inference:

1. Reflexivity: $\alpha \vdash \alpha$ for every $\alpha \in \mathcal{L}$
2. Cut: if $\alpha \vdash \beta \wedge \gamma$ and $\gamma \wedge \delta \vdash \eta$ then $\alpha \wedge \delta \vdash \beta \wedge \eta$

Reflexivity is relatively straightforward, but *Cut* may be harder to decipher at first glance. In essence, *Cut* is a form of logical transitivity, and is the same reasoning as using a lemma for a theorem: anything that can be derived in the current theory, can be used to derive more inferences. Using the above expression for Cut, if α expresses the current information available, and γ is a direct consequence, then observe how since η is a consequence of γ in the presence of δ, then to derive η from α, all that is needed is δ.

While the above two properties are viewed as basic properties a meaningful consequence relation should satisfy, they can both be extended by specifying additional properties, or be discarded, if the form of reasoning desired is not modelled well using those properties. A consequence relation not satisfying reflexivity, for example, may express some form of reasoning about trust, including rejecting information given to the agent by an untrustworthy source [3].

In this way, properties essentially constrain the set of consequence relations of interest, by first assuming the set of all consequence relations to be of interest, and then eliminating those relations not satisfying the desired properties.

2.5 Properties of Classical Deduction

In the context of this dissertation, it is appropriate to discuss a set of properties satisfied by Tarskian notions of entailment. In the previous section, classical entailment, $\models$, was defined semantically. However, the properties classical entailment satisfies can be enumerated in much the same fashion as consequence relations in the previous section. Let $\mathcal{K}$ be a knowledge base, and recall that $\mathcal{C}n$ was defined as a consequence operator on any set of formulas, producing a set of formulas that is inferred, via some notion of consequence, from the set of formulas given as input.

Tarski [67, 56] defined a set of properties that any notion of logical consequence should satisfy:

1. Inclusion: $\mathcal{K} \subseteq \mathcal{C}n(\mathcal{K})$
2. Monotonicity: if $\mathcal{K} \subseteq \mathcal{K}'$ then $\mathcal{C}n(\mathcal{K}) \subseteq \mathcal{C}n(\mathcal{K}')$
3. Idempotence: $\mathcal{C}n(\mathcal{K}) = \mathcal{C}n(\mathcal{C}n(\mathcal{K}))$

Any consequence operation satisfying the above properties is referred to as a Tarskian operation.

Inclusion is reasonably self-evident, a knowledge base entails *at least* the stated explicit information. *Idempotence* states that classical entailment produces a set of formulas that is closed under the consequence operation. In this context, closure states that for some notion of consequence $\models$, a closed set Γ contains every formula α such that $\Gamma \models \alpha$. *Idempotence* therefore ensures that every formula that is a logical consequence of the knowledge base is in the resultant set, and describes a form of reasoning that is omniscient, always producing every possible inference in a single operation.

Monotonicity states that adding any explicit information to a knowledge base should never remove an inference entailed by the knowledge base without said information. At first, *monotonicity* might seem to make perfect sense as a characteristic of reasoning, as adding information should not result in retracting statements if the reasoning process is perfectly sound. For example, consider the knowledge base:

- `man` → `mortal`
- `Socrates` → `man`
- `plucked chicken` → `featherless biped`

Then it is reasonable, just from the first two statements, to conclude that Socrates is mortal. The third statement, while part of the same knowledge base, is irrelevant to our conclusion, and therefore monotonicity ensures that this conclusion is not retracted given new information.

From a semantic perspective, monotonicity essentially states that only those inferences with the maximum support in the knowledge base should be drawn. Recall that for some knowledge base $\mathcal{K}$ and a formula $\alpha \in \mathcal{L}$, then $\mathcal{K} \vDash \alpha$ if and only if α holds in *every* valuation satisfying $\mathcal{K}$. Monotonicity ensures that if there is a single world where $\mathcal{K}$ is true, and α is not, then $\mathcal{K} \nvDash \alpha$. Tarskian consequence therefore insists that every inference must have iron-clad proof of its veracity.

However, monotonicity does not always appear to be an accurate model for how humans reason in real life. Humans frequently make assumptions that supplement their current knowledge, and it is not the case that additional information never changes one's conclusions; it is rather entirely dependent on whether the new information is a) relevant to the conclusion, and b) convincing enough to change one's mind. Monotonicity, however, is completely indifferent to such a nuance. Rather, any such nuance should be handled explicitly in the knowledge base.

Considering again our knowledge base on Socrates, suppose we had:

- $\texttt{man} \rightarrow \texttt{mortal}$
- $\texttt{Socrates} \rightarrow \texttt{man}$
- $\texttt{Socrates} \rightarrow \neg\ \texttt{mortal}$

Now we may not want to draw our previous conclusion that `Socrates` is mortal, in the presence of additional information that directly states his immortality. Examining the valuations for this knowledge base, there are three that are models:
$\{\texttt{man mortal } \overline{\texttt{Socrates}}, \overline{\texttt{man}} \texttt{ mortal } \overline{\texttt{Socrates}}, \overline{\texttt{man}}\ \overline{\texttt{mortal}}\ \overline{\texttt{Socrates}}\}$

In other words, as far as classical entailment is concerned, `Socrates` does not exist, as there is no model of the knowledge base such that `Socrates` is true. For `Socrates` to be able to exist, the knowledge base could be changed such that in place of $\texttt{man} \rightarrow \texttt{mortal}$, instead the statement $\texttt{man} \wedge\neg\ \texttt{Socrates} \rightarrow \texttt{mortal}$ could be added. This would mean the valuation $\texttt{man}\ \overline{\texttt{mortal}}\ \texttt{Socrates}$ would be a model of the knowledge base, and therefore there is a world where `Socrates` exists.

The moral of the above example is that in a monotonic logic, the concept of something *typically* being the case cannot be directly modelled in the language, and the overall intended meaning can be challenging to model. As exceptions have to be explicitly handled in the knowledge base, any domain rich in exceptions, such as flightless birds, will result in unreasonably large formulas, especially as the amount of information to be modelled gets larger and larger. Consider a possible knowledge base about birds containing a formula to the effect of: $\texttt{bird} \land \neg\texttt{penguin} \land \neg\texttt{ostrich} \land \neg\texttt{kiwi} \land \neg\ldots \rightarrow \texttt{flies}$. However, "birds typically fly" is a reasonable statement that accurately expresses reality, as realistically, flight is heavily associated with birds, and the majority of them do indeed fly. Any logic-based system cataloguing birds should likely consider it reasonable to conclude that a random bird does fly in the absence of knowledge to the contrary. Here, the proposed solution to this particular problem is that of nonmonotonic logics.

2.6 Object and Meta Levels

Lastly, before leaving classical logics behind, the notion of the object-level of a language, and the meta-level should be made explicit. In propositional logic, the object-level is any part of the language that is used to model knowledge, i.e., the propositional formulas themselves that make up the language $\mathcal{L}$. The meta-level, on the other hand, is anything that operates over the object-level. Entailment, as mentioned previously, is a meta-level concept, as it is an operation that defines what can be inferred from the object-level. The intuition is that the object-level represents knowledge, and the meta-level is knowledge about that knowledge.

This distinction can be made more obvious by considering the concepts and symbols defined so far. Recall when $\mathcal{L}$ was defined. The connectives used to make up the formulas in $\mathcal{L}$, that is $\neg$, $\land$, $\lor$, $\rightarrow$, $\leftrightarrow$, $\top$, $\bot$ along with the propositional atoms, are all object-level connectives. On the other hand, entailment was defined with the $\vDash$ symbol, and a knowledge base was denoted as $\mathcal{K}$. Both of these symbols are meta-level symbols, that are used to represent what can be said about the object-level statements.

This distinction is not trivial. It is easy to mix up object-level and meta-level concepts, and doing so can result in errors can prove difficult to diagnose. Consider the binary relation for consequence relations, $\vdash$, which represents the notion that a formula is a logical consequence of another formula, or a set of formulas. This is, inherently, a meta-level notion of consequence, $\alpha \vdash \beta$ is not a statement in the language, rather it is a statement about the language, and often is conditional on what form of reasoning $\vdash$ is meant to represent. Contrast this with $\rightarrow$, which is a connective in $\mathcal{L}$, and therefore is an object-level operation. However, even though $\rightarrow$ is associated with a type of notion of consequence, it is fixed, with a truth table defining its meaning in all cases. $\alpha \rightarrow \beta$ states that β must be true in any valuation where α is true, and therefore $\rightarrow$ asserts that this consequence is the case, and has to hold in any meaningful interpretation of the object-level knowledge. Confusing $\alpha \vdash \beta$ for an object-level statement is an easy mistake to make, and can result in adding a meta-level inference to explicit object-level knowledge. Whether or not $\alpha \vdash \beta$ is true or not is completely dependent on the meaning attached to $\vdash$, and cannot be asserted in all cases, in contrast to the object-level. This is the case for any meta-level symbol, such as $\models$, $\equiv$, $\Vdash$, and $\vdash$.

The distinction between object-level and meta-level notions will become a significant issue in later chapters. Defeasible consequence will be defined on the meta-level, and then the same symbol will be redefined to be an object-level connective. This will be made explicit, however the shift from meta-level to object-level should be noted, as the intended meaning behind the symbol changes fundamentally.

Chapter 3

Approaches to Defeasible Reasoning

In the previous chapter, the limitations to expressing exceptional knowledge using classical logics was discussed. It then raises the question, what is the best way to alter existing logics so as to accurately and efficiently model information containing exceptions? A number of competing formalisms have been defined, some of which have been shown to be identical in expressivity, from different perspectives. This chapter will briefly describe a number of these frameworks, before motivating the choice of the framework that the rest of this dissertation will be focused on.

For a discussion on the uses of nonmonotonic reasoning, section 3 from McCarthy [53] enumerates a number of problems and applications for nonmonotonic logics.

3.1 Belief Revision

One popular approach to reasoning about belief change, a problem that corresponds to the core of nonmonotonic reasoning, is belief revision, first defined by Alchourron, Gärdenfors and Makinson (AGM) [1, 2]. Belief revision models an agent's set of beliefs about the world by encoding them as a set of statements in a logic, referred to as a *belief set*, and defines operations that model the agent adjusting the belief set to conform to new information. The two main operations to achieve this are revision and contraction. The revision operator models being told that a given statement, often not in the belief set, is true, and modifies the belief set such that it incorporates the new statement in a satisfiable way. Contraction is the inverse operation, where a statement is provided with the information that it is not entailed from the knowledge base, and the belief set is modified such that the statement is no longer entailed.

AGM belief revision operators conform to a number of postulates, describing the properties the operators should satisfy. These postulates are not without controversy, as there are extensions to belief revision that revolve around changing the postulates to model different forms of reasoning. Darwiche and Pearl [30] proposed an extension to handle conditional beliefs, statements subject to retraction, by changing the AGM postulates to allow for iterated belief change, meaning the changes to a knowledge base are stored after an operation, which is not the case in AGM style belief revision. Booth [8, 9] has worked on a number of extensions to belief revision, including iterated belief revision, as well as non-prioritised belief revision. Non-prioritised belief revision, also referred to as trust-sensitive belief revision [10, 44], is another extension that does not always accept new information, whereas in AGM belief revision, new information is prioritised over existing information. There have also been developments in merging belief revision with nonmonotonic logics by Casini and Meyer [24], which integrates the belief change operators into a preferential framework, by defining the AGM postulates in a preferential logic.

3.2 Circumscription

Circumscription [52] [53] is another framework for nonmonotonic reasoning. First described in 1980, it was proposed as an extension to first-order logic to address the problem of encoding a form of "common sense" reasoning in existing formal structures. For example, a boat is assumed to work as intended, unless there is something preventing its use, and therefore is an atypical boat. Circumscription treats each predicate, or concept, as able to be normal or abnormal. This allows a fluid formalism that can reason elegantly about poorly defined concepts such as "besides x there is something else atypical about y".

The mechanism used to represent such ideas is a meta-level axiom over a regular, monotonic language that circumscribes a formula in the language such that an instance of a formula can be explicitly abnormal with respect to some characteristic, which allows for easy modelling of nested abnormalities. Using circumscription, it is relatively straightforward to represent in a single knowledge base, for example, "normal animals, defined as not being abnormal in a particular way, do not fly", and also "birds normally fly, unless it is abnormal in a different way than previously defined". This allows circumscription to be quite expressive, allowing for numerous levels of defeasible reasoning. However, this generally means that to get meaningful results, one has to pick which predicates may or may not be atypical, which puts a higher burden on the modelling process to accurately capture the intended meaning. Since circumscription considers atypicality for every predicate, the computational costs of implementing circumscription increase exponentially with the size of the knowledge base. Cadoli and Lenzerini [23] showed an algorithm for computing circumscription in the propositional case that takes exponential time in the worst case, and polynomial time in the best case.

3.3 Default Logic

Reiter's default logic is a nonmonotonic framework that has proven influential in the field [48, 36, 31]. Default logic represents information as defaults with the intended meaning of "most x's are y's", or "typically a's are b's". First described by Reiter [61], and then revised by Reiter and Criscuolo [62]. It was originally devised to enrich first-order logic, by adding the notion that there are default states, assumptions that can be drawn as inferences in the absence of information to the contrary, as a solution to the same core problem of nonmonotonic reasoning: how to most effectively model information containing exceptions. Default logic does so by choosing to address a problem arising from modelling around the exceptions: the problem of inheritance for non-exceptional subclasses. If one models birds, and chooses to explicitly handle exceptions by stating "birds that are neither penguins nor ostriches fly", then upon learning that tweety is a bird, but it is not clear which bird, classical entailment will not infer that tweety can fly. Reiter's solution was to therefore add the notion that, by default, it should be assumed that generic birds should fly. A default theory, analogous to a knowledge base for default logic, is a pair $\langle W, D\rangle$ where W is a set of ordinary formulas in the language, and D is a set of default rules: rules of inference of the form:

$$\frac{\alpha(x) : M\beta_1(x)...M\beta_n(x)}{\gamma(x)}$$

This rule can be read as "if x is an α, and it is consistent with all knowledge so far that x is also $\beta_1 ... \beta_n$, then conclude that x is also a γ. Consistency in this context is taken to mean that the negation is not derivable using classical entailment. That is, a formula β is consistent with a default theory Δ if $\Delta \not\models \neg\beta$. Default logic then defines *extensions* to a default theory. Extensions of a default theory are expected to satisfy certain properties: they must contain W, it must be closed under classical deduction, and it should satisfy every default rule in D. Such extensions can be thought of as satisfying different patterns of nonmonotonic reasoning, and can be used to find defeasible inferences of a given default theory.

Reiter defined default logic proof-theoretically, but did not have a corresponding semantics. The lack of a model theory means that it can be difficult to choose between different extensions, having to instead rely on intuition about what kind of reasoning is suitable for a given domain [64]. However, Delgrande et al. [31] defined a semantics for default logic, along with a number of extensions. In section 5.5.1 it will be shown that the general patterns of default reasoning can be formalized semantically using the KLM framework [48].

3.4 Propositional Typicality Logic

In Propositional Typicality Logic (PTL) [11], defeasibility is encoded by a unary operator $\bullet$ that is placed before a formula in any part of a propositional formula. A propositional formula $\alpha \rightarrow \beta$ can be defeasibly expressed as either $\bullet\alpha \rightarrow \beta$ which semantically means that typical αs imply β, as $\alpha \rightarrow \bullet\beta$ which reads that α implies typical βs, $\bullet\alpha \rightarrow \bullet\beta$ meaning that typical αs imply typical βs, or even as $\bullet(\alpha \rightarrow \beta)$, meaning that typically, α implies β. The $\bullet$ operator can, as was just shown, be placed before any formula in the language: the language of PTL can therefore be defined as every formula $\alpha \in \mathcal{L}$, along with $\bullet\alpha$ for every $\alpha \in \mathcal{L}$. The flexibility of this typicality operator allows for a variety of representations of defeasible knowledge.

The semantics of PTL is preferential, meaning that preference relations are defined over valuations, such that there are valuations that are *minimal* with respect to the relation. Therefore, there are no preferred valuations, and the $\bullet$ operator is understood to mean that only those minimal valuations that are models of the formula are to be considered to prove the truth of the statement.

The typicality operator is strong enough that PTL can embed propositional AGM belief revision [11]. Both belief revision and PTL have a preferential semantics, and so revising a knowledge base with some formula α in belief revision is semantically similar to the models of $\bullet\alpha$ in PTL.

3.5 Nonmonotonic Modal Logic

Modal logic is an extension of some logic, here propositional logic will be the language to be extended, that adds two modal operators, usually denoted as $\Box$, and $\Diamond$. Modal operators are unary operations, that can be placed before any formula in the language, that is for any $\alpha \in \mathcal{L}$, $\Box\alpha$ is a valid statement in a modal logic. Given that the modal logic contains negation, $\neg$, then the two operators are interdefinable: given a formula α, and the modal operators $\Box$ and $\Diamond$, then $\Box\alpha \equiv \neg \Diamond \neg\alpha$, and vice versa. Modal operators may also be denoted L, K, or M dependent on the exact meaning ascribed to it in the language in question. For example, in deontic logic, which is intended to reason about duty and normativity, the modal operator $\Box$ carries the interpretation of "it is obligatory" [68], whereas in epistemic logic it could mean "it is known", or "it is common knowledge" [40], and denoted K. Using deontic logic as an example, if $\Box$ is taken to mean "it is obligatory", then using the above equality, $\Diamond$ is used as shorthand for "it is permitted", as it is defined as $\neg \Box \neg\alpha$, for any $\alpha \in \mathcal{L}$, which is read as "it is not obligatory to not α".

In general, modal operators are intended to represent some mode of truth regarding a statement in the language. Rather than examining the absolute truth of a formula, as in propositional logic, modal logic examines the conditions under which some statement is true. Therefore, when a formula is combined with a modal operator, the truth value is not dependent purely on the truth of the formula, but rather on whether the formula *could* become true, or whether it is *always* true, with the interpretation of that truth dependent on the specific modal logic in question.

McDermott and Doyle [54] and McDermott [55] proposed a *nonmonotonic* modal logic, where the modal operator, denoted in his work as M, is read as "is consistent with the knowledge so far". On the meta-level, new rules for nonmonotonic inference are defined. Mp is the case in any theory $\mathcal{K}$ (read: knowledge base) if $\mathcal{K} \vdash \neg p$ is not provable. Naturally, this definition leads to some apparent circular reasoning: what does provable mean? It also allows for spurious entailments, as pointed out by McDermott [55]. McDermott [55] defines a modified semantics based on modal logic semantics: a class of models referred to as *noncommittal* models. Noncommittal models are a subset of all modal models, excluding those models that have undesirable, unfounded necessary statements of the form $\neg M\alpha$ for some α, which results in a set of models that "have as many things possible as possible". The resulting semantics is also accompanied by a proof system, for which soundness and completeness have been proven [55].

However, this form of nonmonotonic modal logic has some theoretical problems. The first formalization by McDermott and Doyle [54] is predicated on $M\alpha$ being the case if α is consistent, but McDermott [55] noted that, as a result of the definition of consistency, $M\alpha$ is not inconsistent with $\neg\alpha$, meaning that there exist knowledge bases in this framework for which it is both the case that $M\alpha$ and $\neg\alpha$ to be derivable, i.e., that it is both the case that α is consistent and that it is false. The second formalization [55] then based consistency on existing modal logic semantics, which resulted in the subsequent nonmonotonic logic collapsing into regular, monotonic modal logic. Moore [58] pointed out these flaws, and used it as a basis for defining autoepistemic logic.

3.6 Autoepistemic Logic

Autoepistemic logic was first defined by Moore [58, 57], building upon the ideas of nonmonotonic modal logic [54, 55]. Epistemic logic, in general, is a family of modal logics that was devised to reason about knowing agents. Correspondingly, autoepistemic logic is, conceptually, a logic representing an agent that reflects on their own beliefs about the world, represented by a set of formulas referred to as an autoepistemic theory. The modal operator, L, then has the interpretation of "it is believed". The modelling of an agent's beliefs necessarily has to allow for nonmonotonic inferences, as it has to account for an agent's beliefs and knowledge change, which may well result in retractions of conclusions. Autoepistemic logic is defined via a model theory: initially, the semantics of which are propositional semantics with an additional layer of an autoepistemic interpretation, which is a propositional model where $L\alpha$ is true if and only if α is true in the autoepistemic theory. Moore [57] defined an alternative model theory for autoepistemic logic more in line with traditional modal logic semantics.

Konolige [45] showed that autoepistemic logic and default logic have an equivalence, that is every default logic theory can be expressed in autoepistemic logic, and vice versa, every statement possible in autoepistemic logic can be expressed as a default theory. One significant implication of this result is that autoepistemic semantics could be ported to default logic, which was defined as a proof-theoretic system, with no model theory.

3.7 Preferential Approach

The approach to nonmonotonic reasoning explored in this dissertation is that of the preferential approach, which was first defined by Shoham [64, 65]. Shoham defined the class of preferential logics, achieved by enriching regular semantics with a preference relation over all valuations. Then, for a formula α to be satisfiable, it must be satisfied by all the most preferred models of α. This set of most preferred models is defined as being all models of α that are minimal with respect to the preference relation. Using the preferential semantics, notions of preferential entailment can be defined by specifying that $\{\alpha\} \mathrel{|\!\!\approx} \beta$, read as "$\alpha$ preferentially entails β", if and only if all preferred models of α also satisfy β. A foundation for a corresponding proof-theoretic system for preferential reasoning was also defined by Gabbay [34]. Both the semantics and corresponding proof system were then extended by Kraus et al. [46], to form what will be referred to as the KLM framework. This exact framework will be described, along with extensions, in the next two chapters.

The main reasons for choosing the KLM framework are the following: it has both a model theory, based on the preferential semantics just mentioned, a proof theory based on that of Gabbay's that also functions as an extendable set of postulates, analogous to that of belief revision's, and it also defines reasoning algorithms that are computationally well behaved and no less efficient than classical reasoning algorithms [49].

Chapter 4

Preferential Reasoning

So far, it has been shown that classical monotonic logics lack the expressivity to explicitly represent exceptions, without challenging alterations to the formulas in the knowledge base. However, this raises the question of what such a *nonmonotonic* logic looks like, and how it should behave. Kraus, Lehmann, and Magidor, who will be referred to from here as KLM [46] argued that a nonmonotonic logic should be able to explicitly state "an x is typically a y", in which they define "typically" to be read as "in the normal case, it is reasonable to conclude y, given x". Classical logic cannot syntactically convey such a reading, and so KLM [46] and Lehmann and Magidor [49] define an extension to propositional logic to capture it. Initially, KLM defined a *preferential consequence relation* over a propositional logic. In other words, they first describe what it means to reason in a nonmonotonic fashion over the regular propositional language. Later, this will be extended to incorporating nonmonotonic notions on the object-level.

This chapter will describe a fundamental type of reasoning in the KLM framework, preferential reasoning. Preferential reasoning is described as the core of nonmonotonic reasoning [26], for reasons that should become clear, but partly because it lays down the semantic foundations that are shared by many nonmonotonic logics. First, preferential consequence relations will be described, along with the semantics associated with them, before defining preferential entailment and preferential logic.

4.1 Preferential Consequence Relations

KLM [46] first defined a meta-level consequence relation on a propositional language. Initially, this was formalized in a series of logics defined by different notions of consequence, each satisfying a more inclusive set of properties. The particular system of interest here was denoted **P**, for preferential reasoning, and will be what is described in this section. A *preferential* consequence relation is denoted by $\mid\!\sim$, to be contrasted with $\vdash$, and is a set of *defeasible implications*, also referred to in the literature as conditional assertions, written as $\alpha \mid\!\sim \beta$ with $\alpha, \beta \in \mathcal{L}$, with the intended reading that "from α, I am willing to jump to conclude β unless I have information to the contrary". In this context, "defeasible" could be read as meaning "retractable", that is any defeasible statement is one that may be withdrawn upon learning contradictory information. As in the classical case, when $\mid\!\sim$ represents a consequence relation, then $\alpha \mid\!\sim \beta$ means that the pair (α, β) is in the consequence relation and $\alpha \not\mid\!\sim \beta$ means that it is not in the relation.

Preferential consequence relations are defined by a set of properties, represented as a set of rules of inference, which should be read as: from the presence of the statements above, the statements below can be derived. In the context of a consequence relation, however, these rules should be interpreted as: if the statements above are in the relation, then the statements below must also be in the relation. The following definition presents these rules:

Definition 4.1. *The consequence relation defined by $\mid\!\sim$ is a preferential consequence relation if and only if it satisfies the following properties, referred to from now as the KLM postulates:*

1. (LLE) Left logical equivalence: $\frac{\top \models \alpha \leftrightarrow \beta,\ \alpha \mid\sim \gamma}{\beta \mid\sim \gamma}$
2. (RW) Right weakening: $\frac{\top \models \alpha \rightarrow \beta,\ \gamma \mid\sim \alpha}{\gamma \mid\sim \beta}$
3. (Ref) Reflexivity: $\alpha \mid\!\sim \alpha$
4. And: $\frac{\alpha \mid\sim \beta,\ \alpha \mid\sim \gamma}{\alpha \mid\sim \beta \wedge \gamma}$
5. Or: $\frac{\alpha \mid\sim \gamma,\ \beta \mid\sim \gamma}{\alpha \vee \beta \mid\sim \gamma}$

6. (CM) Cautious Monotonicity: $\dfrac{\alpha \mid\!\sim \gamma,\ \alpha \mid\!\sim \beta}{\alpha \wedge \beta \mid\!\sim \gamma}$

The above set of properties essentially acts as a set of constraints on all consequence relations. Recall that consequence relations are sets of ordered pairs of the form $\{(\alpha,\beta), (\gamma,\delta), ...\}$ with $\alpha,\beta,\gamma,\delta \in \mathcal{L}$. The above definition states that only those consequence relations, $\mid\!\sim$, satisfying all of the above KLM postulates may be referred to as a preferential consequence relation. A consequence relation satisfies one of these given properties if, given the presence of pairs satisfying the statements above, then there must be a pair, or pairs, satisfying the statements below. For example, *reflexivity* requires that a preferential consequence relation must have, for every $\alpha \in \mathcal{L}$, the pair (α,α).

Each postulate corresponds to a pattern of reasoning about defeasible information. *Left logical equivalence* states that if α and β are classically equivalent, and γ is typically derivable from one of them, then it should be typically derivable from either of them. This rule should be intuitively verifiable, but enforces the influence of the underlying logic on the preferential consequence relation.

Right weakening, as with *LLE*, enforces the influence of the underlying classical logic on the preferential consequence relation. *RW* states that any classical consequences derivable from a defeasible consequence, can themselves be defeasibly concluded from the initial premises. *RW* essentially states that defeasible consequences are closed under logical consequences [49].

Reflexivity is satisfied by almost all types of reasoning, and is an axiom of classical consequence relations. Simply, *reflexivity* requires any preferential consequence relation to enforce all formulas to be defeasible consequences of themselves.

And expresses that the conjunction of two defeasible consequences is itself a defeasible consequence. Consider the statements, isomorphic or identical to the intuitions given by KLM [46]: "If it rains, then usually the streets get wet" and "If it rains, then usually the plants are watered". Then, it should be reasonable to be able to derive "If it rains, then usually the streets get wet, and usually the plants are watered".

Or expresses that if a formula can be defeasibly concluded from two different formulas, then it should be a defeasible conclusion from the disjunction of those formulas. Paraphrasing the example given by KLM [46]: consider that if it is known that "If Andy attends the party, then normally the evening will be great", and that "If Jessie attends the party, then normally the evening will be great", then if it is known that "Andy or Jessie will attend the party", that it should be reasonable to conclude that attending is a good idea.

Finally, *cautious monotonicity* essentially states that upon learning new information, if that information could have been inferred before learning it, then it should never invalidate any conclusion previously derived. If it is known that "it is raining", and it can be inferred that "the roads are wet", and it can also be inferred that "plants are watered", then neither inference should have any bearing on the other, as they are derivable from the same knowledge. *Cautious monotonicity* is a central property of preferential consequence, as it replaces the property of monotonicity previously described as a central property for classical reasoning. Whereas monotonicity states that adding premises does not affect prior conclusions, cautious monotonicity rather states that only premises that are themselves conclusions from the same knowledge are irrelevant to prior conclusions. Therefore, there is room left for premises, that are not already able to be inferred and therefore represents new knowledge, to cause prior conclusions to be withdrawn, and so defeasibility is introduced.

Some other properties of interest that are derivable from the KLM postulates are the following [46]:

1. Cut: $\frac{\alpha \land \beta \mid\!\sim \gamma,\ \alpha \mid\!\sim \beta}{\alpha \mid\!\sim \gamma}$

2. S: $\frac{\alpha \land \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \beta \to \gamma}$

3. D: $\frac{\alpha \land \neg \beta \mid\!\sim \gamma,\ \alpha \land \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma}$

Cut is used in monotonic logics, compare the *cut* rule for classical consequence relations defined previously. Note, however, that in the form presented here it does not imply monotonicity, whereas the usual form of *cut* does.

S is another derived rule, specifically, it is implied by *reflexivity*, *right weakening*, *left logical equivalence*, and *Or*. It essentially states that the conjunction of two formulas defeasibly implying a third implies that one of the conjuncts defeasibly implies a material link between the other conjunct and the conclusion.

The last derived rule of interest, D, is a formalism of proof by cases, also known as proof by exhaustion. It states that if a defeasible consequence holds in complementary cases, then it should hold in isolation.

Two properties of interest not satisfied by preferential consequence relations, as they both imply monotonicity when added to the KLM postulates, are:

- Transitivity: $\frac{\alpha \mid\!\sim \beta, \beta \mid\!\sim \gamma}{\alpha \mid\!\sim \gamma}$
- Contraposition: $\frac{\alpha \mid\!\sim \beta}{\neg\beta \mid\!\sim \neg\alpha}$

Transitivity and contraposition may seem like intuitive properties for $\mid\!\sim$ to satisfy. In the case of transitivity, it necessarily implies monotonicity because the nature of defeasible information is that such inheritance chains can be broken. The nature of nonmonotonicity is that if a penguin is typically a bird and birds typically fly, then it is not necessarily the case that penguins typically fly. With respect to transitivity's representation above: what if α is an atypical example of β, and it is explicitly known that $\alpha \not\mid\!\sim \gamma$? Transitivity therefore cannot hold, as it is antithetical to defeasible reasoning patterns. Contraposition also violates the nature of nonmonotonicity: continuing the previous theme, if birds typically fly, then are things that do not fly typically not birds? It is perhaps debatable, but the possibility of atypical, flightless birds means that a conservative pattern of nonmonotonic reasoning will conclude not, and knowing that something is flightless is not enough information to conclude that they are not a bird.

Preferential consequence relations form a mathematical basis for how preferential reasoning should behave. However, to define what it means for a defeasible inference to follow from a knowledge base, a semantics that corresponds to this form of reasoning needs to be defined. The next section will define the preferential semantics that can be used to further study this particular type of reasoning.

4.2 Preferential Interpretations

The semantics that KLM [46] and Lehmann and Magidor [49] provided for $\mid\!\sim$ is based on the preferential semantics defined previously by Shoham [64, 65, 66]. A preferential semantics imposes an ordering over valuations such that if an agent prefers a valuation u to another valuation v, then the agent will likely consider u before considering v. The notion of an agent preferring one valuation over another can be interpreted in more than one way. One intended meaning is that more *typical* valuations are preferred. A valuation in which birds fly is preferred to one where they do not. It could also be interpreted via a goal oriented lens, where an agent prefers those valuations more aligned with an end goal.

Generally, in the context of defeasible reasoning, preferred worlds will be those that are more normal, or typical, and those worlds that are more improbable will not be considered, unless explicit information forces them to be.

This preference ordering is achieved in preferential semantics by introducing a meta notion of *states*. Each state is mapped to a classical valuation. States are necessarily distinct from valuations, as multiple states may map to the same valuation, and therefore there may be infinitely many states in a given interpretation. So, formally [46, 49]:

Definition 4.2. *A preferential interpretation $\mathcal{P}$ is a triple $\langle S, l, \prec \rangle$ with S being a possibly infinite set of states, $l{:}S \mapsto \mathcal{U}$ is a function mapping states to valuations, and $\prec$ is a strict partial order on S.*

For every $\alpha \in \mathcal{L}$ and some $\mathcal{P}$, let $[\![\alpha]\!]^{\mathcal{P}} \coloneqq \{s \in S, S \in \mathcal{P}, l(s) \Vdash \alpha\}$, in other words, let $[\![\alpha]\!]^{\mathcal{P}}$ be the set of all states in a given preferential interpretation such that the valuation associated with each state satisfies α.

There are some technical properties of preferential interpretations that should be discussed. First, minimality in $\mathcal{P}$ is defined as follows:

Definition 4.3. *For any $\mathcal{P} = \langle S, l, \prec \rangle$, a state $s \in S$ is minimal in $\mathcal{P}$ if and only if there is no $s' \in S$ such that $s' \prec s$. The set of all such s in some $\mathcal{P}$ is referred to as* $\min_{\prec}(\mathcal{P}) = \{s \mid \text{there is no } s' \text{ such that } s' \prec s\}$.

Now, two important properties of preferential interpretations are as follows:

Definition 4.4. *$\mathcal{P}$ is well-founded if and only if $\langle S, \prec \rangle$ is well-founded. $\langle S, \prec \rangle$ is well-founded if and only if for any subset $S' \subseteq S$ there is a $s \in S'$ that is minimal.*

Along with well-foundedness, there is a related property of smoothness:

Definition 4.5. *The partial order $\prec$ for some $\mathcal{P}$ is smooth if and only if for any formula $\alpha \in \mathcal{L}$, it is the case that $[\![\alpha]\!]^{\mathcal{P}}$ has a minimal state s.*

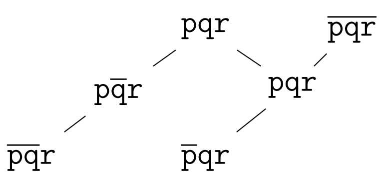
Any preferential interpretation will be said to be finite if and only if the set of states S is finite. Note that smoothness is satisfied by any well-founded preferential interpretation, as well as by any finite interpretation. Well-foundedness essentially prevents an infinitely descending chain of states, while smoothness is a weaker form of well-foundedness that only requires all formulas to have a minimal state that satisfies it.

The significance of preferential interpretations is that each one defines a preferential consequence relation [46]. This can be formalized as follows:

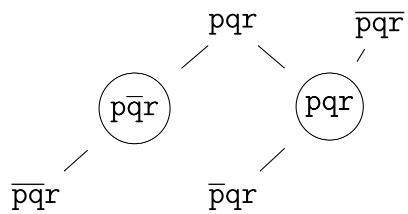
Definition 4.6 (Soundness)**.** *Given a preferential interpretation $\mathcal{P} = \langle S, l, \prec \rangle$ and $\alpha, \beta \in \mathcal{L}$, then $\mathcal{P}$ defines a preferential consequence relation, $\mid\!\sim_{\mathcal{P}}$, such that: $\alpha \mid\!\sim_{\mathcal{P}} \beta$ if and only if for any s minimal in $[\![\alpha]\!]^{\mathcal{P}}$, $s \Vdash \beta$.*

The above soundness definition states that $\mathcal{P}$ defines a preferential consequence relation, $\mid\!\sim_{\mathcal{P}}$, and that $\alpha \mid\!\sim_{\mathcal{P}} \beta$ if in every preferred state where α is true, β is also true. This mechanism, only considering the preferred worlds, essentially provides a conditional view on truth: ignore those scenarios that are improbable in some respect, and focus on those ones that are typical to evaluate whether some β *typically* follows from some α.

To illustrate how a preferential interpretation defines a preferential consequence relation, given the propositional logic over the set of propositions $P \coloneqq \{p, q, r\}$, then the following preferential interpretation, $\mathcal{P}$, can be constructed, where each state is labelled by $l(s)$, the valuation it maps to, rather than the state itself, for readability, and preference is conveyed by a state being underneath and linked to a state to which it is preferred:



Then, by soundness, $\mathcal{P}$ defines a corresponding preferential consequence, $\mid\!\sim_{\mathcal{P}}$, such that for some $\alpha, \beta \in \mathcal{L}$, if for every minimal state satisfying α, β is also satisfied, then $\alpha \mid\!\sim_{\mathcal{P}} \beta$. Then, given the above $\mathcal{P}$, note that for every minimal state satisfying p, that r is also satisfied. The minimal states in question are circled below:

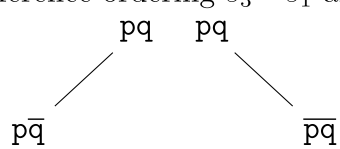


Therefore, $p \mid\!\sim_{\mathcal{P}} r$. There are, of course, other consequences in $\mathcal{P}$, such as $r \mid\!\sim_{\mathcal{P}} \neg p$, as the minimal r states, those that are labelled $\overline{\texttt{pq}}\texttt{r}$ and $\overline{\texttt{p}}\texttt{qr}$, both satisfy $\neg p$.

Using the soundness definition, the following lemma shows why duplicate states, multiple states that map onto the same valuation, are important for preferential interpretations, specifically for the following representation theorem:

Lemma 4.0.1. *There exists a preferential consequence relation such that there does not exist a corresponding preferential interpretation without duplicate states.*

Proof: KLM [46] suggested the following preferential interpretation as an example of such a preferential consequence relation, over the propositional logic with the set of atoms $P = \{p, q\}$: $\mathcal{P} \coloneqq \{S, l, \prec\}$ with $S = \{s_1, s_2, s_3, s_4\}$, the labelling function: $l(s_1) = \texttt{pq}$, $l(s_2) = \texttt{pq}$, $l(s_3) = \overline{\texttt{p}}\texttt{q}$, $l(s_1) = \overline{\texttt{pq}}$, and the preference ordering $s_3 \prec s_1$ and $s_4 \prec s_2$. $\mathcal{P}$ can be visualized as follows:



Then, since $\mathcal{P}$ is a preferential interpretation, it defines a preferential consequence relation $\mid\!\sim_{\mathcal{P}}$. The question is whether there exists a preferential interpretation $\mathcal{P}'$ without duplicate states, such that $\mid\!\sim_{\mathcal{P}'}$ is equivalent to $\mid\!\sim_{\mathcal{P}}$. So, $\mathcal{P}' \coloneqq \{S', l', \prec'\}$ such that $S' = \{s'_1, s'_2, s'_3\}$ as there cannot be more states than valuations, by design, and with the labelling function l such that $l(s'_1) = \texttt{pq}$, $l(s'_2) = \texttt{p}\overline{\texttt{q}}$, $l(s'_3) = \overline{\texttt{pq}}$. Then, the core factor is does there exist a preference relation for these states such that every statement of the form $\alpha \mid\!\sim_{\mathcal{P}} \beta$ satisfied by $\mathcal{P}$ is also satisfied by $\mathcal{P}'$? Firstly, note that $\top \mid\!\sim_{\mathcal{P}} \neg q$ and therefore s'_2 and s'_3 have to be minimal in $\mathcal{P}'$. Then, it is also the case that neither $p \mid\!\sim_{\mathcal{P}} q$ nor $p \mid\!\sim_{\mathcal{P}} \neg q$, and so therefore s'_1 must also be minimal in $\mathcal{P}'$ since s'_2 is minimal. But, if s'_1 is minimal in $\mathcal{P}'$ then $\top \not\mid\!\sim_{\mathcal{P}'} \neg q$ even though $\top \mid\!\sim_{\mathcal{P}} \neg q$. So s'_1 cannot be minimal in $\mathcal{P}'$, although it has already been established that it must be. This is a contradiction, and therefore $\mathcal{P}'$ is impossible to construct, and $\mathcal{P}$ defines a preferential consequence relation that cannot be represented by a preferential interpretation without duplicate states. □

The following representation theorem solidifies the link between preferential interpretations and preferential consequence relations [46, 49]:

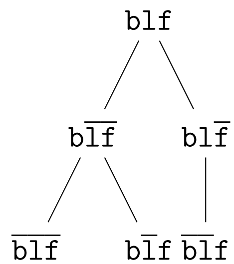
Theorem 4.1. $\mid\!\sim$ *defines a preferential consequence relation if and only if it is the consequence relation defined by a preferential interpretation. In a finite language, every preferential consequence relation is defined by a finite preferential interpretation.*

This representation theorem explicitly links preferential interpretations and preferential consequence relations. Essentially, the theorem shows that every preferential interpretation defines a preferential consequence relation, and vice versa: that every preferential consequence relation is defined by a corresponding preferential interpretation. Preferential interpretations therefore exactly correspond to preferential consequence relations.

To understand the expressivity of preferential interpretations, and associated preferential consequence relations, consider the following defeasible implications:

$$\texttt{boat} \mid\!\sim \texttt{floats} \qquad \texttt{leaky} \mid\!\sim \texttt{boat} \qquad \texttt{leaky} \mid\!\sim \neg\texttt{floats} \tag{4.1}$$

Now, equipped with a preferential semantics, it is possible to evaluate what expressivity has been gained. Recall that the classical semantics will claim that leaky boats do not exist. What if it was certain that they do? Armed with a preferential semantics, it is possible to construct an interpretation in which leaky boats not only exist, but behave as expected. Define a preferential interpretation, $\mathcal{P} = \langle S, l, \prec \rangle$, such that $l(s_1) = \texttt{blf}$, $l(s_2) = \texttt{b}\overline{\texttt{lf}}$, $l(s_3) = \overline{\texttt{blf}}$, $l(s_4) = \texttt{b}\overline{\texttt{l}}\texttt{f}$, $l(s_5) = \texttt{bl}\overline{\texttt{f}}$, $l(s_6) = \overline{\texttt{bl}}\texttt{f}$. $\mathcal{P}$ is depicted visually below, with each node representing a state, such that if $s \prec s'$ then s is visually below s', connected by an edge. For readability, each state is labelled as $l(s)$, i.e., with the valuation associated with the state:



Using $\mathcal{P}$, it is possible to test the above set of statements for satisfiability. For the first statement, it holds that `boat` $\mid\!\sim_{\mathcal{P}}$ `floats` as the minimal state of `boat`, $s_4 : \texttt{b}\overline{\texttt{l}}\texttt{f}$, is also a model of `floats`, and so it is the case that $\min(\llbracket \texttt{boat} \rrbracket^{\mathcal{P}}) \Vdash \texttt{floats}$. For the second statement, the minimal state satisfying `leaky` is $s_5 : \texttt{bl}\overline{\texttt{f}}$, which also satisfies `boat`, and therefore it is the case that `leaky` $\mid\!\sim_{\mathcal{P}}$ `boat`. Lastly, it is the case that `leaky` $\mid\!\sim_{\mathcal{P}} \neg$`floats`, as the minimal state satisfying `leaky`, $s_5 : \texttt{bl}\overline{\texttt{f}}$, is not a model of `floats`. Therefore, every statement in this set of defeasible implications is satisfied by a preferential interpretation, and so there exists a preferential consequence relation, $\mid\!\sim_{\mathcal{P}}$, per the representation theorem defined above, that includes the above statements about leaky boats. Furthermore, $\mathcal{P}$ can be examined for more defeasible implications. For example, note that `float` $\mid\!\sim_{\mathcal{P}} \neg$`leaky` that things that float are typically not leaky, which is an intuitive implication.

Preferential consequence relations and the preferential semantics behind them have been described so far. In the next section, a specific subset of preferential interpretations will be defined.

4.3 Ranked Interpretations

Ranked interpretations are a subset of preferential interpretations for which the partial order satisfies the following property [49]:

Lemma 4.1.1. *If $\prec$ is a strict partial order on a set $\mathcal{V}$, then the following conditions are equivalent:*

1. *for any $x, y, z \in \mathcal{V}$, if $x \nprec y$, $y \nprec x$ and $z \prec x$, then $z \prec y$;*
2. *for any $x, y, z \in \mathcal{V}$, if $x \prec y$, then either $z \prec y$ or $x \prec z$;*
3. *there is a totally ordered set Ω, the strict order denoted by $<$, and a function $r : \mathcal{V} \mapsto \Omega$ called the ranking function such that $s \prec t$ if and only if $r(s) < r(t)$*

Any partial order satisfying the above conditions is referred to as *modular*, and results in a preferential interpretation, that rather than best resembling a graph, instead can be visualized as a number of tiers populated by states, and is referred to as a ranked interpretation due to this key difference. This key difference, between modular partial orders and non-modular partial orders, is simply that any two incomparable states will share the same tier: this is the effect of the properties 1-3 in the above lemma. Given any two states in a ranked interpretation, then either one is preferred, or they have the same rank. This is different to preferential interpretations with a non-modular partial order, where there could be many ways in which two states are incomparable.

There are some interesting properties of ranked interpretations not shared by preferential interpretations. Recall that different states in a preferential interpretation may map to the same valuation. However, if the preference ordering for a preferential interpretation is modular, then if there are two states, s_1 and s_2, in a ranked interpretation, $\mathcal{R}$, that map to the same valuation, there are three possibilities:

1. $s_1 \prec s_2$. In this case, the minimal state is s_1, and s_2 will never be considered. Assuming that this is the only duplicate, then the ranked interpretation $\mathcal{R}$ can be replaced by the ranked interpretation $\mathcal{R}'$ that is equivalent to $\mathcal{R}$ without s_2, and represents the exact same consequence relation as $\mathcal{R}$.

2. s_1 and s_2 have the same rank, and in this case either can be removed without losing or gaining any inferences, and, exactly as in the previous case, can be replaced by an equivalent ranked interpretation that only has one of these states.

3. $s_2 \prec s_1$. This case is equivalent to the first case, and, analogously, s_1 can be removed, creating a new ranked interpretation representing the exact same consequence relation, with fewer states.

The key result of the above is that ranked interpretations do not need to allow duplicate states, as preferential interpretations do. This can be formalized by defining the set of ranked interpretations of interest as being all those ranked interpretations containing no duplicate states. A direct implication of having no duplicate states is that states no longer need to be part of the definition of a ranked interpretation, rather the preference ordering simply can be defined over the set of valuations $\mathcal{U}$.

To illustrate, given the valuations $u, v, w, x, y \in \mathcal{U}$, then two ranked interpretations, $\mathcal{R}_1, \mathcal{R}_2$ can be constructed such that $\mathcal{R}_1 \coloneqq \{\{s_1, s_2, s_3, s_4, s_5\}, l, \prec\}$ with the following mapping from states to valuations:

- $l(s_1) = u$
- $l(s_2) = v$
- $l(s_3) = w$
- $l(s_4) = x$
- $l(s_5) = y$

and such that $\mathcal{R}_2 \coloneqq \{\{s_1, s_2, s_3, s_4, s_5, s_6\}, l', \prec'\}$ with the following mapping from states to valuations:

- $l(s_1) = u$
- $l(s_2) = v$
- $l(s_3) = w$
- $l(s_4) = x$
- $l(s_5) = y$
- $l(s_6) = v$

Then the visual constructions of $\mathcal{R}_1$ and $\mathcal{R}_2$, showing the preference orderings for each are as follows. First $\mathcal{R}_1$:

3	s_5
2	s_4
1	s_3
0	s_1 s_2

And $\mathcal{R}_2$:

3	s_5 s_6
2	s_4
1	s_3
0	s_1 s_2

Notice that in $\mathcal{R}_2$, $l'(s_2) = l'(s_6)$, however $s_2 \prec s_6$. In this case, will s_6 ever affect the answer to a given query? Since only the minimal valuations will be considered, then clearly it will always be overruled by s_2, and be rendered irrelevant. A direct result is that $\mathcal{R}_1$ and $\mathcal{R}_2$ define the exact same consequence relation, and are essentially interchangeable, which implies that the notion of having states, while important for preferential interpretations, is unnecessary for ranked interpretations. Rather, ranked interpretations can be defined as a preference ordering on a set of valuations directly. Since the preference ordering is modular, then, per the third point in Lemma 4.1.1, a ranked interpretation can instead be thought of as a function from the set of valuations to a totally ordered set. Therefore, a ranked interpretation $\mathcal{R}$ can be defined as follows [26]:

Definition 4.7. *A ranked interpretation is a function $\mathcal{R} : \mathcal{U} \mapsto \mathcal{N} \cup \{\infty\}$, satisfying the following convexity property: for every $i \in \mathcal{N}$, if there exists a $u \in \mathcal{U}$ such that $\mathcal{R}(u) = i$, then there must be a $v \in \mathcal{U}$ such that $\mathcal{R}(v) = j$ with $0 \leq j < i$.*

The notation $\mathcal{R}(u)$ will be used to refer to the rank of $u \in \mathcal{U}$ in $\mathcal{R}$. The intuition behind the ranks is that valuations with a lower rank are more typical, or normal, and valuations with infinite rank are impossible. As an example, for some logic over the set of propositions $P \coloneqq \{p, q, r\}$, a ranked interpretation $\mathcal{R}$ can be represented as:

∞	$\mathtt{\overline{pq}r}$ $\mathtt{\overline{p}q\overline{r}}$
2	$\mathtt{\overline{p}qr}$ $\mathtt{p\overline{q}r}$
1	$\mathtt{pqr}$ $\mathtt{pq\overline{r}}$ $\mathtt{p\overline{qr}}$
0	$\mathtt{\overline{pqr}}$

The above shows that $\mathcal{R}(\mathtt{\overline{pq}r}) = \mathcal{R}(\mathtt{\overline{p}q\overline{r}}) = \infty$ and therefore $\mathtt{\overline{pq}r}$ and $\mathtt{\overline{p}q\overline{r}}$ are impossible, and $\mathtt{\overline{pqr}}$ is the most preferred, as $\mathcal{R}(\mathtt{\overline{pqr}}) = 0$, and therefore represents the most typical world. Furthermore, $\mathcal{R}(\mathtt{\overline{p}qr}) = \mathcal{R}(\mathtt{p\overline{q}r}) = 2$, and $\mathcal{R}(\mathtt{pqr}) = \mathcal{R}(\mathtt{pq\overline{r}}) = \mathcal{R}(\mathtt{p\overline{qr}}) = 1$. This notation and particular formulation of ranked interpretations will be preferred for the rest of this dissertation.

Ranked interpretations are also preferential interpretations, and as such they inherit all the properties of preferential interpretations. This includes that just as a preferential interpretation defines a consequence relation, so do ranked interpretations. This includes the notion of minimal states in preferential interpretations, which shall be redefined as follows following the above definition of ranked interpretations:

Definition 4.8. *Given a ranked interpretation $\mathcal{R}$ and any formula $\alpha \in \mathcal{L}$, it holds that $u \in [\![\alpha]\!]^{\mathcal{R}}$ is minimal if and only if there is no $v \in [\![\alpha]\!]^{\mathcal{R}}$ such that $\mathcal{R}(v) < \mathcal{R}(u)$*

However, ranked interpretations do not exactly define just a preferential consequence relation, rather they define a *rational* consequence relation, which is a specific type of preferential consequence relation, per the following definition [49]:

Definition 4.9. *A rational consequence relation is a consequence relation, $\mid\!\sim$, that satisfies the previous KLM properties: Ref, And, Or, LLE, RW, CM, along with the following property:*

$$\text{Rational monotonicity (RM): } \frac{\alpha \mid\!\sim \gamma,\ \alpha \not\mid\!\sim \neg\beta}{\alpha \wedge \beta \mid\!\sim \gamma}$$

Rational consequence relations are linked to ranked interpretations in the following way, analogous to preferential relations [49]:

Lemma 4.1.2. *If $\mathcal{R}$ is a ranked interpretation, then it defines a rational consequence relation $\mid\!\sim_{\mathcal{R}}$ such that $\alpha \mid\!\sim_{\mathcal{R}} \beta$ if and only if for any u minimal in $[\![\alpha]\!]^{\mathcal{R}}$, $u \Vdash \beta$.*

Given the same $\mathcal{R}$ as above, then this lemma states that the consequence relation it generates, $\mid\!\sim_{\mathcal{R}}$ is rational, and is defined with respect to $\mathcal{R}$ such that for all $\alpha, \beta \in \mathcal{L}$ then $\alpha \mid\!\sim_{\mathcal{R}} \beta$ if and only if all minimal valuations satisfying α also satisfy β. Note the valuations circled in $\mathcal{R}$ below:

∞	$\overline{\mathtt{pq}}\mathtt{r}$ $\overline{\mathtt{p}}\mathtt{q}\overline{\mathtt{r}}$
2	$\overline{\mathtt{p}}\mathtt{qr}$ $\mathtt{p}\overline{\mathtt{q}}\mathtt{r}$
1	(pqr) ($\mathtt{pq}\overline{\mathtt{r}}$) $\mathtt{p}\overline{\mathtt{qr}}$
0	$\overline{\mathtt{pqr}}$

Both of the above circled valuations, pqr and $\mathtt{pq}\overline{\mathtt{r}}$, are the minimal valuations of the atom q. Also note that both of them satisfy the atom p, and so therefore $\mathtt{q} \mid\!\sim_{\mathcal{R}} \mathtt{p}$.

The importance of a consequence relation being rational lies in comparing *cautious monotonicity* and *rational monotonicity*. In essence, *CM* states that only information already derivable using the current knowledge will never cause a retraction of an inference. On the other hand, *rational monotonicity* weakens this constraint, but strengthens what can be included in the consequence relation, by requiring that any new information, the negation of which is *not* derivable using the current knowledge, should never cause a retraction. The key difference is while *cautious monotonicity* states that any knowledge already able to be inferred never invalidates a conclusion, *rational monotonicity* states that any new knowledge that does not contradict with any prior inferences should not invalidate a conclusion. This is a fundamental shift in reasoning, as *rational monotonicity*, in contrast to any other postulate, represents a *negative* rule. Whereas the other rules are phrased "in the presence of this information, this is derivable", *RM* is phrased as "in the *absence* of this information, this is derivable". The significance of this is that it describes a model of reasoning that is much more willing to draw a speculative inference that does not directly contradict any existing information. Especially compared to cautious monotonicity; it should be noted that given rational monotonicity, cautious monotonicity is implied, and therefore not technically necessary when discussing rational consequence.

Why is *rational monotonicity* desirable? Consider the defeasible implication $\alpha \mid\!\sim \beta$ with α and β being different propositional formulas. Let γ be a formula in $\mathcal{L}$. Then it is intuitive to expect $\alpha \land \gamma \mid\!\sim \beta$ to be inferred. The reasoning has been extensively discussed [46, 49] and boils down to that unless explicitly stated, there is nothing that should be assumed about γ that would influence α to no longer imply β, and it is therefore sensible to infer that $\alpha \land \gamma$ statements are not so different to α statements. This reasoning, however, is not enforced by *cautious monotonicity*. Observe that there is no reason to necessarily infer γ, and so therefore $\alpha \land \gamma \mid\!\sim \beta$ cannot be inferred by a preferential consequence relation. However, note that *rational monotonicity* will infer such a statement, as $\neg\gamma$ is not inferred in the presence of $\alpha \mid\!\sim \beta$, and so *rational monotonicity* will derive $\alpha \land \gamma \mid\!\sim \beta$. This is a strong argument for *rational monotonicity* as a nonmonotonic replacement for monotonicity.

The next theorem is a representation theorem for rational consequence relations and ranked interpretations [49]:

Theorem 4.2. *A consequence relation* $\mid\!\sim$ *on* $\mathcal{L}$ *is rational if and only if it is the consequence relation defined by a ranked interpretation. Given a finite* *language, every rational consequence relation is defined by a finite ranked interpretation.*

The above representation theorem completes the link started by Lemma 4.1.2., by confirming it as an if and only if.

Ranked interpretations are a significant subset of preferential interpretations, however, in the next section preferential semantics will be revisited, to define a notion of entailment with respect to these semantics.

4.4 Preferential Entailment

The question as to how to define entailment with respect to a preferential semantics, is asking how to define what defeasible inferences follow from a set of defeasible information. This represents a shift in the syntax. Note that until now, $\mid\sim$ was treated as a meta-level consequence relation over a classical propositional language. Now that an entailment relation based on preferential semantics is to be defined, $\mid\sim$ will now be treated as an object-level connective, specifying explicitly defeasible information. This can be formalized by defining a new language: $\mathcal{L}_P \coloneqq \mathcal{L} \cup \{\alpha \mid\sim \beta \mid \alpha, \beta \in \mathcal{L}\}$. From now, the language referred to in this dissertation will be that of $\mathcal{L}_P$. Intuitively, $\mathcal{L}_P$ is that of propositional logic with the added connective $\mid\sim$, which is a defeasible counterpart to $\rightarrow$ and may be read as "typically implies", but that may not be nested. No formula of the type $(\alpha \mid\sim \beta) \mid\sim \gamma$ is valid in $\mathcal{L}_P$. A formal definition of a defeasible implication is then:

Definition 4.10. *A defeasible implication is a statement $\alpha \mid\sim \beta \in \mathcal{L}_P$ where $\alpha, \beta \in \mathcal{L}$.*

A set of defeasible implications forms a defeasible knowledge base:

Definition 4.11. *A defeasible knowledge base, $\mathcal{K}$, is a set of defeasible implications, $\alpha \mid\sim \beta$.*

In this context, a defeasible knowledge base, also referred to in the literature as a conditional knowledge base, represents the *explicit* information, as opposed to the set of inferences that follows from it. The semantics of $\mathcal{L}_P$ is that of the previously defined preferential semantics, and satisfaction in a preferential interpretation can be defined as follows:

Definition 4.12. *Given a preferential interpretation $\mathcal{P}$ and a defeasible implication $\alpha \mid\sim \beta$, $\mathcal{P}$ satisfies $\alpha \mid\sim \beta$, written $\mathcal{P} \Vdash \alpha \mid\sim \beta$ if and only if for every s minimal in $[\![\alpha]\!]^{\mathcal{P}}$, $s \Vdash \beta$. If $\mathcal{P} \Vdash \alpha \mid\sim \beta$ then $\mathcal{P}$ is said to be a model of $\alpha \mid\sim \beta$.*

The above definition can be extended to sets of defeasible implications:

Definition 4.13. *Given a defeasible knowledge base $\mathcal{K}$, a preferential interpretation $\mathcal{P}$ satisfies $\mathcal{K}$, written $\mathcal{P} \Vdash \mathcal{K}$ if and only if for every $\alpha \mid\sim \beta \in \mathcal{K}$, $\mathcal{P} \Vdash \alpha \mid\sim \beta$. If $\mathcal{P} \mid\sim \mathcal{K}$, then $\mathcal{P}$ is said to be a model of $\mathcal{K}$.*

A defeasible implication can also be a classical formula, as any classical propositional formula can be expressed as a defeasible implication in the following way:

Definition 4.14. *A preferential interpretation $\mathcal{P}$ satisfies a formula $\alpha \in \mathcal{L}$, written $\mathcal{P} \Vdash \alpha$, if and only if for all states $s \in \mathcal{P}$ then $s \Vdash \alpha$.*

The above definition states that a classical formula is satisfied in a preferential interpretation if it is satisfied by every state in the preferential interpretation. This allows us to define classical formulas as a defeasible implication:

Corollary 4.2.1. *Any formula $\alpha \in \mathcal{L}$ can be expressed as a defeasible implication $\neg\alpha \mid\sim \bot$. For any preferential interpretation $\mathcal{P}$, $\mathcal{P} \Vdash \alpha$ if and only if $\mathcal{P} \Vdash \neg\alpha \mid\sim \bot$.*

Note that $[\![\bot]\!] = \varnothing$, by definition, as $\bot$ is the propositional constant representing *false*. Then, since $\mathcal{P} \Vdash \neg\alpha \mid\sim \bot$ if and only if $[\![\neg\alpha]\!]^{\mathcal{P}} \subseteq \varnothing$, there are no minimal states where α is false, and therefore α is true in $\mathcal{P}$. Therefore, any mention of a defeasible implication may also refer to classical statements, and the nonmonotonic logic $\mathcal{L}_P$ contains the classical propositional logic $\mathcal{L}$, and, in fact, extends classical logic [26]. Then, any result regarding defeasible implications also applies to classical statements, by noting that any defeasible implication $\alpha \mid\sim \beta$ is equivalent to a classical statement γ whenever α is logically equivalent to $\neg\gamma$ and β is logically equivalent to $\bot$.

Just as in classical logic, entailment, denoted with $\models$, is a meta-level reasoning concept; from now any type of meta-level defeasible entailment is denoted by $\mid\approx$. Then, the statement $\mathcal{K} \mid\approx \alpha \mid\sim \beta$ should be read as "the knowledge base defeasibly entails that α typically implies β". Recall that a query is a propositional formula. Queries now will refer to any defeasible implication $\alpha \mid\sim \beta$, which includes any classical statement α.

This section will define the first entailment relation that uses preferential semantics by applying Tarskian notions of consequence [56, 67], called preferential entailment [49].

Definition 4.15. *Given a defeasible knowledge base* $\mathcal{K}$*, and a defeasible implication* $\alpha \mid\!\sim \beta$*, preferential entailment, denoted* $\approx_P$*, is defined as:* $\mathcal{K} \approx_P \alpha \mid\!\sim \beta$ *if and only if for every preferential model,* $\mathcal{P}$*, of* $\mathcal{K}$*,* $\mathcal{P} \Vdash \alpha \mid\!\sim \beta$*.*

Preferential entailment essentially works analogously to classical entailment: if a query is satisfied by every preferential model of a knowledge base, then it is preferentially entailed by the knowledge base. This parallels classical entailment semantics, adjusted for preferential semantics.

The following theorems from KLM [46] and Lehmann and Magidor [49] are significant in characterizing preferential entailment.

Theorem 4.3. *Given a defeasible knowledge base* $\mathcal{K}$*, and a defeasible implication* $\alpha \mid\!\sim \beta$*, the following conditions are equivalent:*

1. $\alpha \mid\!\sim \beta$ *is preferentially entailed by* $\mathcal{K}$*, i.e.,* $\mathcal{K} \approx_P \alpha \mid\!\sim \beta$*.*

2. *for all preferential consequence relations* $\mid\!\sim_{\mathcal{P}}$*, defined by a preferential interpretation* $\mathcal{P}$*, if* $\mathcal{P} \Vdash \mathcal{K}$*, then it is the case that* $\alpha \mid\!\sim_{\mathcal{P}} \beta$*.*

3. $\alpha \mid\!\sim \beta$ *has a proof from* $\mathcal{K}$ *using the following KLM postulates:*

   - *(LLE) Left logical equivalence:* $\frac{\mathcal{K} \approx \alpha \leftrightarrow \beta, \mathcal{K} \approx \alpha \mid\sim \gamma}{\mathcal{K} \approx \beta \mid\sim \gamma}$
   - *(RW) Right weakening:* $\frac{\mathcal{K} \approx \alpha \rightarrow \beta, \mathcal{K} \approx \gamma \mid\sim \alpha}{\mathcal{K} \approx \gamma \mid\sim \beta}$
   - *(Ref) Reflexivity:* $\mathcal{K} \approx \alpha \mid\!\sim \alpha$
   - *And:* $\frac{\mathcal{K} \approx \alpha \mid\sim \beta, \mathcal{K} \approx \alpha \mid\sim \gamma}{\mathcal{K} \approx \alpha \mid\sim \beta \wedge \gamma}$
   - *Or:* $\frac{\mathcal{K} \approx \alpha \mid\sim \gamma, \mathcal{K} \approx \beta \mid\sim \gamma}{\mathcal{K} \approx \alpha \vee \beta \mid\sim \gamma}$
   - *(CM) Cautious Monotonicity:* $\frac{\mathcal{K} \approx \alpha \mid\sim \gamma, \mathcal{K} \approx \alpha \mid\sim \beta}{\mathcal{K} \approx \alpha \wedge \beta \mid\sim \gamma}$

4. $\alpha \mid\!\sim \beta \in \mathcal{C}n^{\mathcal{P}}(\mathcal{K})$*, i.e.,* $\alpha \mid\!\sim \beta$ *is in the consequence operation corresponding to preferential entailment,* $\mathcal{C}n^{\mathcal{P}}$*, of* $\mathcal{K}$*.*

Point (2) defines preferential entailment in terms of preferential consequence relations. Initially, Lehmann and Magidor [49] defined preferential entailment in these terms. Essentially, it is possible to characterize the pattern of reasoning of preferential entailment as a consequence relation: given a defeasible knowledge base $\mathcal{K}$, then define a consequence relation $\mid\sim$ such that for any $\alpha, \beta \in \mathcal{L}$, then $\alpha \mid\sim \beta$ if for every preferential interpretation $\mathcal{P}$ that is a model of $\mathcal{K}$, it holds that $\alpha \mid\sim_{\mathcal{P}} \beta$. The resulting consequence relation $\mid\sim$ is preferential and has a single corresponding preferential interpretation. Alternatively, the consequence relation corresponding to preferential entailment can be viewed as the intersection of all preferential consequence relations: the set of defeasible implications agreed upon by all preferential consequence relations.

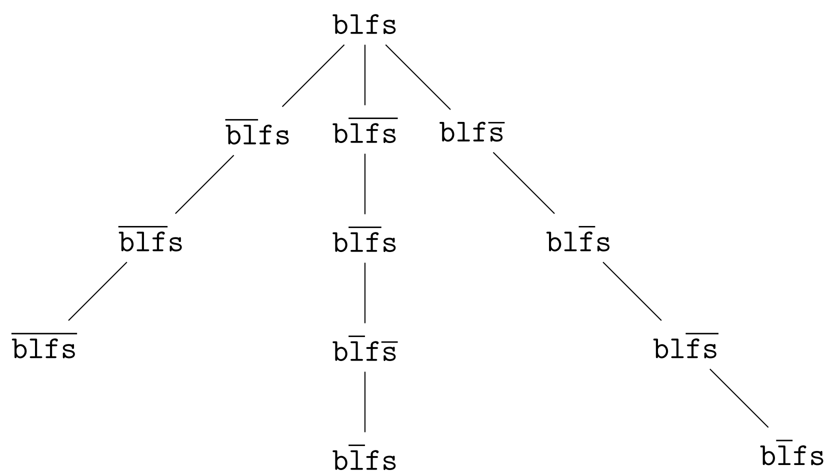
Point (3) of the above theorem rephrases the KLM postulates that previously were defined as properties satisfied by all preferential consequence relations. Recall that the properties previously were constraints on a consequence relation $\mid\sim$, and as such $\mid\sim$ was a meta-level notion of consequence. Now, as $\mid\sim$ is an object-level connective, specifying explicit information, they are presented slightly differently. $\mid\approx$ is now the meta-level notion of consequence, and $\mid\sim$ is an object-level connective. In this form, the KLM postulates form a deductive system for preferential entailment, where the set of inference rules, R, are the KLM postulates as written above, and, in fact, forms a Gentzen system for a defeasible language. This means that, given some $\mathcal{K}$, applying the KLM postulates as inference rules will yield all $\alpha \mid\sim \beta$ that are preferentially entailed by $\mathcal{K}$. Alternatively, every $\alpha \mid\sim \beta$ preferentially entailed by $\mathcal{K}$ has a proof from $\mathcal{K}$ using some combination of the KLM postulates. Naturally, this means that preferential entailment satisfies the KLM postulates, and therefore, by the representation theorem for preferential interpretations, it can be defined by a single preferential interpretation, which will be formalised in the following corollary of the above theorem, also described by Lehmann and Magidor [49]:

Corollary 4.3.1. *The set $\mathcal{C}n^{\mathcal{P}}(\mathcal{K})$ is a preferential consequence relation, and therefore there is a preferential model satisfying exactly the defeasible implications of $\mathcal{C}n^{\mathcal{P}}(\mathcal{K})$. If $\mathcal{K}$ is a preferential consequence relation, then $\mathcal{K} = \mathcal{C}n^{\mathcal{P}}(\mathcal{K})$. $\mathcal{C}n^{\mathcal{P}}(\mathcal{K})$ grows monotonically with $\mathcal{K}$.*

The above corollary, from Lehmann and Magidor [49], confirms that there is a single preferential interpretation that can be used to define preferential entailment, via the representation theorem defined in the previous section. Since there is a single preferential interpretation, $\mathcal{P}$, that will infer the same set of defeasible implications as preferential entailment, given the same knowledge base $\mathcal{K}$, it then follows that there is a preferential consequence relation, $\mid\!\sim_{\mathcal{P}}$, that will also infer the same set of defeasible implications as preferential entailment.

The other significant implication of the last corollary is that preferential entailment is still monotonic. Given a defeasible knowledge base $\mathcal{K}$, if $\mathcal{K} \mid\!\approx_{\mathcal{P}} \alpha \mid\!\sim \beta$, then for any defeasible implication $\gamma \mid\!\sim \delta$, it will always hold that $\mathcal{K} \cup \{\gamma \mid\!\sim \delta\} \mid\!\approx_{\mathcal{P}} \alpha \mid\!\sim \beta$, and so if nonmonotonic entailment is the goal, then the search continues.

By way of example, consider the knowledge base $\mathcal{K} \coloneqq \{\mathtt{boat} \mid\!\sim \mathtt{floats}, \mathtt{boat} \mid\!\sim \mathtt{sailors}, \neg(\mathtt{leaky} \to \mathtt{boat}) \mid\!\sim \bot, \mathtt{leaky} \mid\!\sim \neg\mathtt{floats}\}$, containing similar statements as from a previous section, with some additional defeasible implications. Consider the following query: $\mathcal{K} \mid\!\approx_{P} \mathtt{leaky} \mid\!\sim \mathtt{sailors}$, or do leaky boats typically have sailors? Recall that for a defeasible implication to be preferentially entailed, it must be satisfied by every preferential model of the knowledge base. Consider the following preferential interpretation, $\mathcal{P}^{\mathcal{K}} = \langle S, l, \prec \rangle$ with $S = \{l(s_1) = \mathtt{blfs}, l(s_2) = \mathtt{blf\overline{s}}, l(s_3) = \mathtt{\overline{bl}fs}, l(s_4) = \mathtt{b\overline{lfs}}, l(s_5) = \mathtt{bl\overline{f}s}, l(s_6) = \mathtt{bl\overline{fs}}, l(s_7) = \mathtt{b\overline{l}fs}, l(s_8) = \mathtt{\overline{blf}s}, l(s_9) = \mathtt{\overline{blfs}}, l(s_{10}) = \mathtt{b\overline{lf}s}, l(s_{11}) = \mathtt{b\overline{l}f\overline{s}}, l(s_{12}) = \mathtt{b\overline{l}fs}\}$. Note that, for brevity, each valuation refers to each proposition by the first letter only. The following tree representation of $\mathcal{P}^{\mathcal{K}}$, as before, uses the valuations, $l(s)$ as node labels, for readability:



Is the above preferential interpretation a model of $\mathcal{K}$? For it to be so, each statement in $\mathcal{K}$ must be satisfied in the above interpretation. The first statement, $\mathtt{boat} \mid\!\sim \mathtt{floats}$ is a defeasible implication, and therefore only the minimal $\mathtt{boat}$ valuations need be checked in $\mathcal{P}^{\mathcal{K}}$. The minimal $\mathtt{boat}$ valuation is $\mathtt{b\overline{l}fs}$, which is also a model of $\mathtt{floats}$, and so $\mathcal{P}^{\mathcal{K}} \Vdash \mathtt{boat} \mid\!\sim \mathtt{floats}$. The second statement $\mathtt{boat} \mid\!\sim \mathtt{sailors}$ can be checked against the same valuation as the prior statement, as again only the minimal $\mathtt{boat}$ valuation need be checked. Again $\mathtt{b\overline{l}fs} \Vdash \mathtt{sailors}$, and so $\mathcal{P}^{\mathcal{K}} \Vdash \mathtt{boat} \mid\!\sim \mathtt{sailors}$. The third statement, $\neg(\mathtt{leaky} \rightarrow \mathtt{boat}) \mid\!\sim \bot$, is logically equivalent to the classical statement $\mathtt{leaky}{\rightarrow}\mathtt{boat}$ and so needs to hold in all valuations of $\mathcal{P}^{\mathcal{K}}$, not just the minimal valuations. All valuations satisfying $\mathtt{leaky}$ are: $\mathtt{blfs}$, $\mathtt{blf\overline{s}}$, $\mathtt{bl\overline{f}s}$, and $\mathtt{bl\overline{fs}}$, all of which are also models of $\mathtt{boat}$, and so therefore $\mathcal{P}^{\mathcal{K}} \Vdash \neg(\mathtt{leaky} \rightarrow \mathtt{boat}) \mid\!\sim \bot$. Lastly, the defeasible statement $\mathtt{leaky} \mid\!\sim \neg\mathtt{floats}$ is satisfied by this interpretation, if and only if all the minimal valuations satisfying $\mathtt{leaky}$ do not satisfy $\mathtt{floats}$. The minimal $\mathtt{leaky}$ valuation is: $\mathtt{bl\overline{fs}}$, which does not satisfy $\mathtt{floats}$, and so therefore it is the case that $\mathcal{P}^{\mathcal{K}} \Vdash \mathtt{leaky} \mid\!\sim \neg\mathtt{floats}$, and with the last statement it can be concluded that $\mathcal{P}^{\mathcal{K}} \Vdash \mathcal{K}$, that this preferential interpretation is a model of $\mathcal{K}$. Recall that the query was whether $\mathcal{K} \mid\!\approx_P \mathtt{leaky} \mid\!\sim \mathtt{sailors}$. Does this interpretation satisfy the query statement? The minimal valuation of $\mathtt{leaky}$ is $\mathtt{bl\overline{fs}}$, which is not a model of $\mathtt{sailors}$, and so therefore $\mathcal{P}^{\mathcal{K}} \nVdash \mathtt{leaky} \mid\!\sim \mathtt{sailors}$. What is the significance of this? Recall that a knowledge base $\mathcal{K}$ preferentially entails a defeasible statement α if and only if every preferential model of $\mathcal{K}$ is also a model of α. The above interpretation is a counterexample, and so it can be concluded that $\mathcal{K} \not\mid\!\approx_P \mathtt{leaky} \mid\!\sim \mathtt{sailors}$. This example is representative of what preferential entailment cannot entail. This pattern of reasoning, inheritance of properties for exceptional sub-classes, is one not satisfied by preferential entailment.

However, there are schools of thought that claim that preferential entailment is enough [39]. Preferential reasoning is perfectly adequate to reason about a defeasible knowledge base, since the object language is nonmonotonic, and $\mid\!\sim$ is strong enough for a nonmonotonic logic. Giordano et al. [39] argue that the addition of rational monotonicity as an inference rule may result in undesirable inferences, and therefore preferential reasoning is strong enough. However, their argument is based in description logic ABox reasoning, and therefore not particularly relevant in the context of propositional logic.

However, this work is, of course, interested in nonmonotonic reasoning, and in that context preferential entailment is, while a significant formalism, still not enough to be an acceptable entailment relation for defeasible reasoning.

4.5 Ranked Entailment

Having defined ranked interpretations and preferential entailment, it is worth examining the same result restricted to ranked interpretations. It was shown that preferential entailment is the result of a Tarskian pattern of entailment using preferential interpretations, and so it is now time to investigate what ranked entailment looks like. It should be clear that the pattern this process will follow will be identical to the process of defining preferential entailment from preferential interpretations.

Therefore, define ranked entailment as follows [49]:

Definition 4.16. *Given a knowledge base $\mathcal{K}$, and a defeasible implication $\alpha \mid\sim \beta$, $\mathcal{K} \approx_{\mathcal{R}} \alpha \mid\sim \beta$, read as $\mathcal{K}$ rank entails $\alpha \mid\sim \beta$, if and only if for every ranked interpretation, $\mathcal{R}$, such that $\mathcal{R} \Vdash \mathcal{K}$, it is the case that $\mathcal{R} \Vdash \alpha \mid\sim \beta$.*

In other words, $\mathcal{K}$ rank entails a defeasible implication $\alpha \mid\sim \beta$ if every ranked interpretation satisfying $\mathcal{K}$ also satisfies $\alpha \mid\sim \beta$, and vice versa. The result is exactly preferential entailment [49]:

Theorem 4.4. *Given a knowledge base $\mathcal{K}$ and a defeasible implication $\alpha \mid\sim \beta$ it is the case that $\mathcal{K} \approx_{\mathcal{R}} \alpha \mid\sim \beta$ if and only if $\mathcal{K} \approx_{P} \alpha \mid\sim \beta$.*

There are a few rationales for ranked entailment being identical to preferential entailment. One explanation is that ranked entailment represents all of those inferences that every ranked interpretation satisfies. Each of these inferences, via the representation theorem, has a proof using the KLM postulates, including *rational monotonicity.* It should be noted that each postulate is phrased positively, meaning that each state "in the presence of the above, derive the below", with the exception of *RM*, which is phrased "in the presence of one statement, *and the absence of another statement* above, derive the statement below". This results in a number of different inferences satisfied by different interpretation. When taking the intersection of all these different interpretations, which results in ranked entailment, the result is the smallest group of inferences that all interpretations agree on. This intuitively would result in the set of inferences with a proof using the first six postulates, and without any inferences with a proof using *RM*, as there would exist at least one interpretation that does not agree with such an inference.

An interesting conclusion that can be drawn from the above, is that rational consequence relations are not closed under intersection [48]. This is different from preferential consequence relations, as it has been demonstrated that the intersection of all preferential relations is itself a preferential relation: preferential entailment. However, the intersection of all rational consequence relations is not a rational relation; rather it is a preferential relation: again preferential entailment.

Since ranked entailment is still monotonic, the question of how to define a nonmonotonic entailment relation using preferential semantics remains. Additionally, as will be described, it is not enough for an entailment relation to be nonmonotonic, as there are very many nonmonotonic entailment relations, in contrast to classical entailment. Rather, nonmonotonic entailment relations that are rational in the same way as rational consequence relations - satisfying the KLM postulates - are the class of entailment relations that are desirable. The intuition behind this is straightforward: given a defeasible knowledge base, $\mathcal{K}$, and a nonmonotonic entailment relation, $\mid\approx$, then the set of statements of the form $\alpha \mid\sim \beta$ that follow from $\mathcal{K}$, along with every $\alpha \mid\sim \beta$ already in $\mathcal{K}$, together form a consequence relation $\mid\sim_{CR}$. If $\mid\approx$ is rational, then $\mid\sim_{CR}$ is rational. The next chapter will describe how to generate such rational nonmonotonic entailment relations.

Chapter 5

Nonmonotonic Reasoning

Defeasible knowledge bases have been defined, as well as preferential and ranked entailments for those knowledge bases, which were shown to still be monotonic. This chapter will describe how to define nonmonotonic entailment relations using preferential semantics. The core entailment relation that will be the focus of the majority of this chapter is *rational closure*. The main reason for this is that it was the first nonmonotonic entailment relation defined by Lehmann and Magidor [49], and, due to its status as the most conservative nonmonotonic entailment relation, forms what will be termed the nonmonotonic core of defeasible entailment, analogous to how preferential entailment will be shown to be the monotonic core. First, rational closure will be defined semantically, and then it will be shown how this semantics can be used to define rational closure over the statements in the knowledge base itself. After rational closure is described, a general framework for generating defeasible entailment will be built up, with the view of defining iterative classes of entailment relations by extending the KLM framework with more properties, isolating desirable entailment relations in a systematic fashion.

5.1 Minimal Ranked Entailment

In the previous chapter, ranked entailment was described and was shown to be exactly preferential entailment, and therefore monotonic. The question then remains, how to define a semantics for a nonmonotonic entailment relation? This section will answer such a question, by describing what will be termed *minimal ranked entailment*.

First, a definition for some notation:

Definition 5.1. *Given a defeasible knowledge base $\mathcal{K}$ and any $\alpha, \beta \in \mathcal{L}$ the notation $\alpha < \beta$ in $\mathcal{K}$ is shorthand for $\mathcal{K} \approx \alpha \vee \beta \mid\!\sim \neg\beta$. Additionally, $\alpha \leq \beta$ is shorthand for $\beta < \alpha$ not being in $\mathcal{K}$.*

Now, consider two consequence operators, each resulting in a different rational consequence relation, $\mathcal{C}n_1$ and $\mathcal{C}n_2$, operating over the same knowledge base $\mathcal{K}$. Lehmann and Magidor [49] showed that it is possible to define a preference ordering between them, such that $\mathcal{C}n_1(\mathcal{K}) \prec \mathcal{C}n_2(\mathcal{K})$ if and only if:

1. there exists a defeasible implication $\alpha \mid\!\sim \beta$ in $\mathcal{C}n_2(\mathcal{K}) \setminus \mathcal{C}n_1(\mathcal{K})$ such that for all γ such that $\gamma < \alpha$ in $\mathcal{C}n_1(\mathcal{K})$, and for all δ such that $\gamma \mid\!\sim \delta$ is in $\mathcal{C}n_1(\mathcal{K})$, there is also $\gamma \mid\!\sim \delta$ in $\mathcal{C}n_2(\mathcal{K})$, and

2. for any λ, θ if $\lambda \mid\!\sim \theta$ is in $\mathcal{C}n_1(\mathcal{K}) \setminus \mathcal{C}n_2(\mathcal{K})$ there is a defeasible implication $\rho \mid\!\sim \eta$ in $\mathcal{C}n_2(\mathcal{K}) \setminus \mathcal{C}n_1(\mathcal{K})$ such that $\rho < \lambda$ for $\mathcal{C}n_2(\mathcal{K})$

The above definition can prove difficult to understand intuitively, however the intuition is that the valid arguments for hypothetical proponents of different rational consequence relations might go as follows:

1. Proponent of $\mathcal{C}n_1$: your relation contains an implication, $\alpha \mid\!\sim \beta$, that mine does not, and therefore contains an unsupported inference.

2. Proponent of $\mathcal{C}n_2$: yes, but your relation contains an implication $\gamma \mid\!\sim \delta$ that mine does not, and you yourself think that γ refers to a situation that is more usual than the one referred to by α.

In the formal definition above, (1) refers to a valid attack that a relation A may have against another relation B, and (2) refers to the absence of such an attack existing for B against A. Using the above intuition, in the formal definition above, $\mathcal{C}n_1(\mathcal{K}) \prec \mathcal{C}n_2(\mathcal{K})$ if $\mathcal{C}n_1(\mathcal{K})$ has an attack that $\mathcal{C}n_2(\mathcal{K})$ cannot defend against, while the negation is true for $\mathcal{C}n_2(\mathcal{K})$, meaning that it has no attack of its own.

The intuitive definition above describes a valid attack, and a corresponding valid defense against an attack, which essentially describes two incomparable notions of consequence. This is not a complete correspondence to the formal definition, which defines when they are comparable: (2) in the formal definition describes the negation of the attack, i.e., that $\mathcal{C}n_2$ has no attack, and not that $\mathcal{C}n_2$ has a defense.

Given the above defined preference relation, a partial order can be defined over all ranked models of $\mathcal{K}$. This partial order, denoted $\preceq_{\mathcal{K}}$ can be defined semantically as follows [26]:

Definition 5.2. *Given a knowledge base, $\mathcal{K}$, and $\mathcal{R}^{\mathcal{K}}$ the set of all ranked interpretations of $\mathcal{K}$, it holds for every $\mathcal{R}_1^{\mathcal{K}}, \mathcal{R}_2^{\mathcal{K}} \in \mathcal{R}^{\mathcal{K}}$ that $\mathcal{R}_1^{\mathcal{K}} \preceq_{\mathcal{K}} \mathcal{R}_2^{\mathcal{K}}$ if and only if for every $u \in \mathcal{U}$, $\mathcal{R}_1^{\mathcal{K}}(u) \leq \mathcal{R}_2^{\mathcal{K}}(u)$.*

The intuition is that ranked interpretations that have their valuations "pushed down" as much as possible are preferred, and such interpretations also represent a more conservative pattern of reasoning.

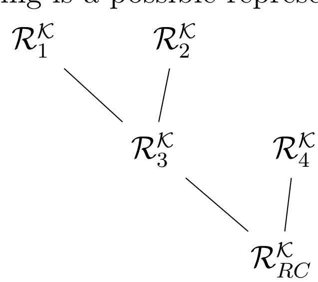
The two different levels of preference orders employed in the above definition should be noted. Observe the distinction between the modular partial order within each ranked interpretation, that defines a total pre-order on valuations, and the more meta-level partial order over all ranked interpretations defining which interpretations have valuations that are more "pushed down", which corresponds to a more conservative form of reasoning.

Giordano et al. [38] showed that the partially ordered set $\langle \mathcal{R}, \preceq_{\mathcal{K}} \rangle$ has a minimal element, which will be denoted as $\mathcal{R}_{RC}^{\mathcal{K}}$. This leads us to the following definition:

Definition 5.3. *Given a defeasible knowledge base $\mathcal{K}$, the minimal ranked interpretation satisfying $\mathcal{K}$, $\mathcal{R}_{RC}^{\mathcal{K}}$, defines an entailment relation, $\mid\!\approx$, called minimal ranked entailment, such that for any defeasible implication $\alpha \mid\!\sim \beta$, $\mathcal{K} \mid\!\approx \alpha \mid\!\sim \beta$ if and only if $\mathcal{R}_{RC}^{\mathcal{K}} \Vdash \alpha \mid\!\sim \beta$.*

That $\mathcal{R}_{RC}$ defines a rational entailment relation is a result of the representation theorem. In the same way that a ranked interpretation can generate a rational consequence relation, similarly it generates an entailment relation satisfying all the KLM postulates.

To show the intuition behind $\preceq_{\mathcal{K}}$, given a defeasible knowledge base $\mathcal{K}$, there will be a finite number of ranked interpretations, $\mathcal{R}^{\mathcal{K}}$. Then, the following is a possible representation of the partially ordered set $\langle \mathcal{R}^{\mathcal{K}}, \preceq_{\mathcal{K}} \rangle$:



The key takeaway from the above visual is that $\mathcal{R}^{\mathcal{K}}_{RC}$ is the minimal element of the set, and will always be for any knowledge base. The structure and size of the set will differ, but there will be a minimum, and that minimum will be designated $\mathcal{R}^{\mathcal{K}}_{RC}$.

Each preferential interpretation can be viewed as codifying a pattern of reasoning, some that are sensible, and others not. Preferential entailment can then be viewed as the pattern of reasoning that accepts all conclusions that are agreed upon by every preferential interpretation, and it was shown that there is a single preferential interpretation that corresponds exactly to preferential entailment's pattern of reasoning. Then, as every ranked interpretation is also a preferential interpretation, each ranked interpretation also codifies a pattern of reasoning. $\mathcal{R}_{RC}$, as the minimal interpretation, represents the most conservative pattern of truly defeasible reasoning codified by a ranked interpretation.

As an example to illustrate that minimal ranked entailment is truly nonmonotonic, consider our leaky boat knowledge base, now with the Flying Dutchman: $\mathcal{K} \coloneqq \{\mathtt{boat} \mid\!\sim \mathtt{floats},\ \neg(\mathtt{leaky} \rightarrow \mathtt{boat}) \mid\!\sim \bot,\ \mathtt{leaky} \mid\!\sim$
$\neg\mathtt{floats}, \neg(\mathtt{FlyingDutchman} \rightarrow \mathtt{boat}) \mid\!\sim \bot\}$

All that is known about the flying dutchman at this point is that it is a boat, and therefore the assumption is made that it is a typical boat. The minimal ranked model of this knowledge base, $\mathcal{R}^{\mathcal{K}}$, is therefore:

∞	$\mathtt{bf\overline{l}d}$ $\mathtt{\overline{b}fld}$ $\mathtt{\overline{bf}ld}$ $\mathtt{\overline{b}fl\overline{d}}$ $\mathtt{\overline{bf}l\overline{d}}$ $\mathtt{\overline{b}fl\overline{d}}$ $\mathtt{\overline{b}f\overline{l}d}$ $\mathtt{\overline{bfl}d}$
2	$\mathtt{bfl\overline{d}}$ $\mathtt{bfld}$
1	$\mathtt{b\overline{f}l\overline{d}}$ $\mathtt{b\overline{f}ld}$
0	$\mathtt{\overline{bfld}}$ $\mathtt{\overline{b}f\overline{ld}}$ $\mathtt{bf\overline{ld}}$ $\mathtt{bf\overline{l}d}$

For brevity, the propositions have been shortened: `b` is `boat`, `f` is `floats`, `l` is `leaky` and `d` is the `FlyingDutchman`. So, what does this ranked model entail? Notice that $\mathcal{R}^{\mathcal{K}} \Vdash \mathtt{FlyingDutchman} \mid\!\sim \mathtt{floats}$, or in other words it follows that the Flying Dutchman typically floats, which makes sense, since it is known that it is a boat, and it is known that boats typically float. Suppose the statement $\neg(\mathtt{FlyingDutchman} \rightarrow \mathtt{leaky}) \mid\!\sim \bot$ was added to the knowledge base. So, the adjusted minimal ranked model of $\mathcal{K} \cup \{\neg(\mathtt{FlyingDutchman} \rightarrow \mathtt{leaky}) \mid\!\sim \bot\}$, which shall be denoted $\mathcal{R}^{\mathcal{K}}_1$ should look like:

∞	$\mathtt{bf\overline{l}d}$ $\mathtt{\overline{b}fld}$ $\mathtt{\overline{bf}ld}$ $\mathtt{\overline{b}fl\overline{d}}$ $\mathtt{\overline{bf}l\overline{d}}$ $\mathtt{\overline{b}fl\overline{d}}$ $\mathtt{\overline{b}f\overline{l}d}$ $\mathtt{\overline{bfl}d}$ $\mathtt{bf\overline{l}d}$
2	$\mathtt{bfl\overline{d}}$ $\mathtt{bfld}$
1	$\mathtt{b\overline{f}l\overline{d}}$ $\mathtt{b\overline{f}ld}$
0	$\mathtt{\overline{bfld}}$ $\mathtt{\overline{b}f\overline{ld}}$ $\mathtt{bf\overline{ld}}$

Now, notice that $\mathcal{R}_1^{\mathcal{K}} \not\Vdash$ `FlyingDutchman` $\mid\!\sim$ `floats`, and instead it is the case that $\mathcal{R}_1^{\mathcal{K}} \Vdash$ `FlyingDutchman` $\mid\!\sim \neg$`floats`. In other words, upon adding information, a previous inference was retracted. This is the nonmonotonic nature of minimal ranked entailment. Suppose this was taken a step further, and the statement $\neg$(`FlyingDutchman` $\rightarrow$ `floats`) $\mid\!\sim \bot$ was added to the knowledge base. The resultant minimal ranked model, denoted $\mathcal{R}_2^{\mathcal{K}}$, should be:

∞	$\mathtt{bf\overline{l}d}$ $\mathtt{\overline{b}fld}$ $\mathtt{\overline{bf}ld}$ $\mathtt{\overline{b}fl\overline{d}}$ $\mathtt{\overline{bf}l\overline{d}}$ $\mathtt{\overline{b}fl\overline{d}}$ $\mathtt{\overline{b}f\overline{l}d}$ $\mathtt{\overline{bfl}d}$ $\mathtt{bf\overline{l}d}$ $\mathtt{b\overline{f}ld}$
2	$\mathtt{bfl\overline{d}}$ $\mathtt{bfld}$
1	$\mathtt{b\overline{f}l\overline{d}}$
0	$\mathtt{\overline{bfld}}$ $\mathtt{\overline{b}f\overline{ld}}$ $\mathtt{bf\overline{ld}}$

Now notice that $\mathcal{R}_2^{\mathcal{K}} \Vdash$ `FlyingDutchman` $\rightarrow$ `floats` and $\mathcal{R}_2^{\mathcal{K}} \Vdash$ `boat` $\wedge$ `leaky` $\mid\!\sim \neg$`floats`. In this way, exceptionality can be "nested" using minimal ranked entailment, with exceptions to exceptions being handled sensibly. The above minimal ranked models of $\mathcal{K}$, can be used to tell the story of an agent that employs a form of common sense reasoning, which still is nevertheless relatively conservative. Initially, all that is known about boats is that they float, with the exception of those that are leaky. Additionally, the agent is aware of a specific boat called the Flying Dutchman, but that is all. From this, the agent is willing to conclude that the Flying Dutchman is a typical boat, that floats. Upon learning that the Flying Dutchman is in fact leaky, then the agent revises his conclusions, and no longer concludes that it is a typical boat, but that it is a typical leaky boat, and therefore does not float. The agent has therefore changed their mind, and retracted a conclusion about the Flying Dutchman being able to float. However, the agent then learns that in fact the Flying Dutchman is a ghost ship, and despite being leaky is still able to float. Now the agent knows that it is not even a typical leaky boat, and withdraws the conclusion that it does not float.

5.2 Rational closure

There is an alternative syntactic definition of minimal ranked entailment, which is a direct result of the semantic definition from the previous section, as well as preferential entailment. First, a necessary preliminary definition from Lehmann and Magidor [49]:

Definition 5.4. *The material counterpart of a defeasible implication* $\alpha \mid\!\sim \beta$ *is the propositional formula* $\alpha \rightarrow \beta$*. Given a defeasible knowledge base* $\mathcal{K}$*, the material counterpart of* $\mathcal{K}$*, denoted* $\overrightarrow{\mathcal{K}}$*, is the set of material counterparts,* $\alpha \rightarrow \beta$*, for every defeasible implication* $\alpha \mid\!\sim \beta \in \mathcal{K}$*.*

The material counterpart is also referred to as a *materialisation*. The utility of the materialisation of a knowledge base was also demonstrated by Lehmann and Magidor [49], as it can be used to calculate those classical formulas satisfied by the most preferred valuations of a defeasible knowledge base, per the following lemma:

Lemma 5.0.1. *Given a defeasible knowledge base* $\mathcal{K}$*, and* $\alpha \in \mathcal{L}$ *a formula, it holds that* $\overrightarrow{\mathcal{K}} \vDash \alpha$ *if and only if* $\mathcal{K} \mid\!\approx_P \top \mid\!\sim \alpha$*.*

The utility of the above lemma may be dubious, at first, but it is a key result for defining exceptionality [49]:

Definition 5.5. *Given a knowledge base,* $\mathcal{K}$*, a propositional formula,* α*, is said to be exceptional for* $\mathcal{K}$ *if and only if* $\mathcal{K} \mid\!\approx_P \top \mid\!\sim \neg\alpha$*.*

In natural language, the above definition states that a formula is exceptional for a knowledge base if the negation of the formula is typical in every preferential interpretation satisfying the knowledge base. Each formula $\alpha \in \mathcal{L}$ is essentially a statement about the world. An exceptional formula can be thought of as a statement about the world that is false in all the best worlds of $\mathcal{K}$, but it may perhaps be true in a less typical world.

Exceptionality uses preferential entailment to demarcate levels of specificity in a knowledge base. Depending on the information encoded, there may be one to many levels of exceptional information.

Then, this can be used to define a function ε on a knowledge base that returns a subset of the knowledge base consisting of every defeasible statement having an antecedent exceptional in the original knowledge base. Formally:

$$\varepsilon(\mathcal{K}) := \{\alpha \mid\!\sim \beta \mid \mathcal{K} \mid\!\approx_P \top \mid\!\sim \neg\alpha\}$$

Using this function, it is possible to define an iterative sequence of knowledge bases as follows:

Definition 5.6. *$\mathcal{E}_0^{\mathcal{K}} \equiv_{def} \mathcal{K}$, $\mathcal{E}_i^{\mathcal{K}} \equiv_{def} \varepsilon(\mathcal{E}_{i-1}^{\mathcal{K}})$ for every $0 < i < n$, and $\mathcal{E}_\infty^{\mathcal{K}} \equiv_{def} \mathcal{E}_n^{\mathcal{K}}$, where n is the smallest k such that $\mathcal{E}_k^{\mathcal{K}} = \mathcal{E}_{k+1}^{\mathcal{K}}$.*

It should be noted that $\mathcal{E}_\infty^{\mathcal{K}} = \varnothing$, under the conditions that there are no formulas in the knowledge base that are exceptional for every level. In the case that $\mathcal{E}_\infty^{\mathcal{K}}$ is non-empty, then it represents those formulas in $\mathcal{K}$ that are logically equivalent to $\bot$, and therefore a classical formula, α, or the defeasible representation $\neg\alpha \mid\!\sim \bot$, is in the rational closure of a knowledge base if and only if $br(\neg\alpha) = \infty$ [26]. Formulas with infinite rank are also referred to as having no rank in the literature [48, 49].

Note that since $\mathcal{K}$ is defined as finite, n must exist. Now that the sequence $\mathcal{E}_0^{\mathcal{K}} ... \mathcal{E}_\infty^{\mathcal{K}}$ is defined, the *base rank* of a formula may be defined [26]:

Definition 5.7. *$br_{\mathcal{K}}(\alpha)$, referred to as the base rank of α, with respect to a given defeasible knowledge base $\mathcal{K}$, is the smallest r such that α is not exceptional in $\mathcal{E}_r^{\mathcal{K}}$, and is defined: $br_{\mathcal{K}}(\alpha) \coloneqq \min\{r \mid \mathcal{E}_r^{\mathcal{K}} \not\approx_R \top \mid\!\sim \neg\alpha\}$*

It should be noted that the base rank of any defeasible implication is the base rank of the antecedent: $br_{\mathcal{K}}(\alpha \mid\!\sim \beta) \equiv br_{\mathcal{K}}(\alpha)$. The intuition behind assigning each defeasible implication a rank is that the lower the rank, the more defeasible the statement. Given two formulas $\alpha, \beta \in \mathcal{L}$, if $br(\alpha) < br(\beta)$, then α represents a more general statement than β, and therefore is more defeasible information, while defeasible implications with infinite rank are classical statements, true in every valuation in every ranked model of $\mathcal{K}$.

Given the knowledge base $\mathcal{K} \coloneqq \{\texttt{boat} \mid\!\sim \texttt{floats}, \texttt{leaky} \land \neg\texttt{boat} \mid\!\sim \bot, \texttt{leaky} \mid\!\sim \neg\texttt{floats}\}$, the above partitioning procedure will provide the following:

- $\mathcal{E}_0 = \mathcal{K} = \{\texttt{boat} \mid\!\sim \texttt{floats}, \texttt{leaky} \land \neg\texttt{boat} \mid\!\sim \bot, \texttt{leaky} \mid\!\sim \neg\texttt{floats}\}$ then it is the case that $\mathcal{E}_0 \approx_P \top \mid\!\sim \neg\texttt{leaky}$ and $\mathcal{E}_0 \approx_P \top \mid\!\sim \neg(\texttt{leaky} \land \neg\texttt{boat})$, and so `leaky` and $\texttt{leaky} \land \neg\texttt{boat}$ are both exceptional, while $\texttt{boat} \mid\!\sim \texttt{floats}$ is not.

- Then, $\mathcal{E}_1 = \{\texttt{leaky} \land \neg\texttt{boat} \mid\!\sim \bot, \texttt{leaky} \mid\!\sim \neg\texttt{floats}\}$, for which it holds that $\mathcal{E}_1 \approx_P \top \mid\!\sim \neg(\texttt{leaky} \land \neg\texttt{boat})$ and so $\texttt{leaky} \land \neg\texttt{boat} \mid\!\sim \bot$ is exceptional for $\mathcal{E}_1$ and $\texttt{leaky} \mid\!\sim \neg\texttt{floats}$ is not exceptional.

- $\mathcal{E}_2 = \{\texttt{leaky} \wedge \neg\texttt{boat} \mid\!\sim \bot\}$ for which $\mathcal{E}_2 \mid\!\approx_P \neg(\texttt{leaky} \wedge \neg\texttt{boat})$ holds, and so therefore $\texttt{leaky} \wedge \neg\texttt{boat} \mid\!\sim \bot$ is exceptional for $\mathcal{E}_2$

- $\mathcal{E}_3 = \mathcal{E}_2 = \{\texttt{leaky} \wedge \neg\texttt{boat} \mid\!\sim \bot\}$ and so the procedure terminates. Therefore, $\texttt{boat} \mid\!\sim \texttt{floats}$ has base rank 0, $\texttt{leaky} \mid\!\sim \neg\texttt{floats}$ has base rank 1, and $\texttt{leaky} \wedge \neg\texttt{boat} \mid\!\sim \bot$ has base rank ∞.

Using this definition of base rank, a form of nonmonotonic entailment called *rational closure* [49] can be defined as such [38]:

Definition 5.8. *Given a knowledge base $\mathcal{K}$, a defeasible implication $\alpha \mid\!\sim \beta$ is said to be in the rational closure of $\mathcal{K}$, and written $\mathcal{K} \mid\!\approx_{RC} \alpha \mid\!\sim \beta$, if and only if $br_{\mathcal{K}}(\alpha) < br_{\mathcal{K}}(\alpha \wedge \neg\beta)$ or $br_{\mathcal{K}}(\alpha) = \infty$.*

The above definition of a defeasible entailment relation uses the base rank of a propositional formula, that acts directly on the syntax. While this is a consequence of previously defined semantics, this particular approach can be implemented purely with a classical entailment SAT solver [49]. At this point, the connection between rational closure and minimal ranked entailment, defined previously, may be of interest. The following observation, also from Giordano et al. [38] provides the fundamental connection between base ranks and the ranks of valuations in the minimal ranked model $\mathcal{R}^{\mathcal{K}}_{RC}$:

Theorem 5.1. *For every knowledge base $\mathcal{K}$ and $\alpha \in \mathcal{L}$, $br_{\mathcal{K}}(\alpha) = \min\{i \mid \exists v \in \hat{\alpha}$ such that $\mathcal{R}^{\mathcal{K}}_{RC}(v) = i\}$.*

A consequence of the above observation is that rational closure will conclude the exact same set of inferences as minimal ranked entailment:

Corollary 5.1.1. *Given a defeasible knowledge base $\mathcal{K}$ and a defeasible implication $\alpha \mid\!\sim \beta$, it holds that $\mathcal{K} \mid\!\approx_{RC} \alpha \mid\!\sim \beta$ if and only if $\mathcal{R}^{\mathcal{K}}_{RC} \Vdash \alpha \mid\!\sim \beta$.*

That is, minimal ranked entailment is in fact the semantic definition of rational closure.

5.3 Algorithm for Rational Closure

In the previous sections, rational closure was defined semantically, and then as base ranks on propositional sentences. Now, an algorithm to compute the rational closure of a defeasible knowledge base $\mathcal{K}$ will be presented.

The first procedure to define, `BaseRank`, is an algorithm on $\mathcal{K}$ that computes the base rank, br, defined in the previous section, of every formula, mapping every explicit propositional formula in $\mathcal{K}$ to the set of natural numbers and infinity: $\{0, 1, 2, 3, ...\} \cup \infty$. `BaseRank` takes a defeasible knowledge base $\mathcal{K}$ as input, but the majority of the procedure is performed on the materialisation, $\overrightarrow{\mathcal{K}}$, recall the definition as $\overrightarrow{\mathcal{K}} \coloneqq \{\alpha \rightarrow \beta \mid \alpha \mathrel{|\!\sim} \beta \in \mathcal{K}\}$. The output will therefore be a tuple of sets of classical implications that are the material counterparts to the defeasible implications in $\mathcal{K}$, corresponding to the sequence $\mathcal{E}_n^{\mathcal{K}}$ of exceptional subsets of $\mathcal{K}$.

As required for a well-defined base rank, the algorithm satisfies the specified constraints for br, that there can be no "gaps" in between ranks, and `BaseRank` will assign each propositional formula the lowest possible rank, while still respecting exceptionality.

Algorithm 1 BaseRank

Input: A knowledge base $\mathcal{K}$
Output: An ordered tuple $(R_0, ..., R_{n-1}, R_\infty, n)$
i := 0;
$E_0 := \overrightarrow{\mathcal{K}}$;
while $E_{i-1} \neq E_i$ **do**
    $E_{i+1} := \{\alpha \rightarrow \beta \in E_i \mid E_i \models \neg\alpha\}$;
    $R_i := E_i \setminus E_{i+1}$;
    $i := i + 1$;
end while
$R_\infty := E_{i-1}$;
if $E_{i-1} = \varnothing$ **then**
    $n := i - 1$;
else
    $n := i$;
end if
return $(R_0, ..., R_{n-1}, R_\infty, n)$

The next algorithm, `RationalClosure`, takes as input $\mathcal{K}$, and a defeasible query $\alpha \mathrel{|\!\sim} \beta$, and returns **true** if the query is entailed by $\mathcal{K}$, and **false** otherwise. `RationalClosure` makes use of a classical entailment checker a polynomial number of times in the size of $\mathcal{K}$, and therefore computing the rational closure is not a computationally harder problem than classical entailment.

Algorithm 2 RationalClosure

Input: A knowledge base $\mathcal{K}$, and a defeasible implication $\alpha \mid\sim \beta$
Output: **true**, if $\mathcal{K} \mid\approx \alpha \mid\sim \beta$, and **false** otherwise
$(R_0, ..., R_{n-1}, R_\infty, n) := \text{BaseRank}(\mathcal{K})$;
$i := 0$
$R := \bigcup_{i=0}^{j<n} R_j$;
while $R_\infty \cup R \models \neg\alpha$ and $R \neq \varnothing$ **do**
    $R := R \setminus R_i$;
    $i := i + 1$;
end while
return $R_\infty \cup R \models \alpha \rightarrow \beta$;

The above algorithm essentially works by checking if there exists an exceptional subset of the knowledge base such that the query is entailed. If the materialisation of the full knowledge base classically entails the negation of the antecedent of the defeasible query, then the antecedent is an exceptional formula, as well as the query itself, and so it removes all statements of the lowest rank in the knowledge base, and performs the same check, until the antecedent is no longer exceptional. Once a rank is found for which the antecedent is not exceptional, then the algorithm checks if the materialisation of the query is entailed, and outputs the result.

It has been shown by Freund [33] that `RationalClosure`, given a defeasible knowledge base $\mathcal{K}$ and a query $\alpha \mid\sim \beta$, returns true if and only if $\alpha \mid\sim \beta$ is in the rational closure of $\mathcal{K}$:

Theorem 5.2. *Given a knowledge base* $\mathcal{K}$ *and a query* $\alpha \mid\sim \beta$ *as input,* `RationalClosure` *returns* ***true*** *if and only if* $\mathcal{K} \mid\approx_{RC} \alpha \mid\sim \beta$.

To illustrate how this algorithm works, consider the knowledge base $\mathcal{K} := \{\texttt{boat} \mid\sim \texttt{floats}, \texttt{leaky} \mid\sim \texttt{boat}, \texttt{leaky} \mid\sim \neg\texttt{floats}, \texttt{FlyingDutchman} \mid\sim \texttt{boat}, \texttt{FlyingDutchman} \mid\sim \texttt{leaky}\}$, and query $\texttt{FlyingDutchman} \mid\sim \neg\texttt{floats}$. Then the material counterpart to $\mathcal{K}$ is $\overrightarrow{\mathcal{K}} := \{\texttt{boat} \rightarrow \texttt{floats}$, $\texttt{leaky} \rightarrow \texttt{boat}$, $\texttt{leaky} \rightarrow \neg\texttt{floats}$, $\texttt{FlyingDutchman} \rightarrow \texttt{boat}$, $\texttt{FlyingDutchman} \rightarrow \texttt{leaky}\}$, and the material counterpart to the query is $\texttt{FlyingDutchman} \rightarrow \neg\texttt{floats}$.

Then applying the `BaseRank` algorithm will assign each formula in $\overrightarrow{\mathcal{K}}$, which now can be designated as $\mathcal{E}_0$, in the following way:

1. $\overrightarrow{\mathcal{K}} = \mathcal{E}_0 \models \neg$`leaky` and $\mathcal{E}_0 \models \neg$`FlyingDutchman` and so $\mathcal{E}_1 \coloneqq \{$`leaky` $\rightarrow$ `boat`, `leaky` $\rightarrow \neg$`floats`, `FlyingDutchman` $\rightarrow$ `boat`, `FlyingDutchman` $\rightarrow$ `leaky`$\}$.

2. $\mathcal{E}_1$ has no exceptional antecedents, and so $\mathcal{E}_2 \coloneqq \varnothing$.

3. $\mathcal{E}_2$ has no exceptional antecedents and so $\mathcal{E}_3 \coloneqq \varnothing$. Since $\mathcal{E}_2 = \mathcal{E}_3$, then the loop is terminated.

The `BaseRank` algorithm will then output the following tuple: $(\{$`boat` $\rightarrow$ `floats`$\}$, $\{$`leaky` $\rightarrow$ `boat`, `leaky` $\rightarrow \neg$`floats`, `FlyingDutchman` $\rightarrow$ `boat`, `FlyingDutchman` $\rightarrow$ `leaky`$\}$, $\varnothing, 3)$. This tuple is then used to construct the following rankings of the formulas in $\overrightarrow{\mathcal{K}}$:

R_∞	$\varnothing$
R_1	`leaky` $\rightarrow$ `boat` `leaky` $\rightarrow \neg$`floats` `FlyingDutchman` $\rightarrow$ `boat` `FlyingDutchman` $\rightarrow$ `leaky`
R_0	`boat` $\rightarrow$ `floats`

Then, to apply the `RationalClosure` algorithm, first it needs to be checked if $R_0 \cup R_1 \cup R_\infty \models \neg$`FlyingDutchman`. Note that it does, as there no models of $R_0 \cup R_1 \cup R_\infty$ that satisfy `FlyingDutchman`. Then, all formulas of rank 0 are removed, leaving the following knowledge base, $\mathcal{E}_1$:

∞	$\varnothing$
1	`leaky` $\rightarrow$ `boat` `leaky` $\rightarrow \neg$`floats` `FlyingDutchman` $\rightarrow$ `boat` `FlyingDutchman` $\rightarrow$ `leaky`

Then, is it the case that $R_1 \cup R_\infty \models \neg$`FlyingDutchman`? Enumerating the set of valuations, written in the same form used previously, satisfying $R_1 \cup R_\infty$ yields the set $\mathcal{U}= \{\mathtt{bl\overline{f}d}, \mathtt{bl\overline{fd}}, \mathtt{b\overline{l}f\overline{d}}, \mathtt{b\overline{lfd}}, \mathtt{\overline{bl}f\overline{d}}, \mathtt{\overline{blfd}}\}$, of which there is a valuation where `FlyingDutchman` is true, and so $R_1 \cup R_\infty \not\models \neg$`FlyingDutchman`. So then `FlyingDutchman` is not exceptional, and now the algorithm checks if $R_1 \cup R_\infty \models$ `FlyingDutchman` $\rightarrow \neg$`floats`. Note that only one valuation, $\mathtt{bl\overline{f}d}$, has `FlyingDutchman` as true, and since `floats` is false in this valuation, it is indeed the case that $R_1 \cup R_\infty \models$ `FlyingDutchman` $\rightarrow \neg$`floats`, and so the output will be **true**.

The above example illustrates how rational closure, a nonmonotonic entailment relation, can be computed algorithmically using a regular classical satisfiability checker. This example can be checked against the example for minimal ranked entailment to see that it does indeed correspond to the rational closure of $\mathcal{K}$.

5.4 Rational Entailment

So far, what has been described is the beginnings of a framework for defeasible reasoning using a preferential semantics, from which some syntactic approaches may be formalised. However, just one entailment relation, rational closure, has been described so far. Rational closure is generated from the minimal ranked model of a knowledge base. Per the representation theorem, every other ranked model of a knowledge base also generates a rational defeasible entailment relation. This raises the question: are all such entailment relations worthy of being examined? The immediate answer is that they are not [26]. The KLM postulates, while elegant and useful, are not quite restrictive enough to isolate the useful ranked models of a knowledge base. They do not, for example, require that every defeasible statement given explicitly in the knowledge base is defeasibly entailed. The KLM postulates were not intended to purely describe all useful defeasible entailment, either, as providing a nonmonotonic deductive system that corresponds to a preferential semantics is a complete system. However, the goal here is to extend it with the view that it is possible, and useful, to create a system such that any common sense pattern of reasoning can be formalised as a defeasible entailment relation, and that only those common sense patterns of reasoning are described.

So, then what is a reasonable way to prune the ranked models of a knowledge base such that only those sensible entailment relations may be generated? Firstly, the class of ranked interpretations examined until this point all satisfy the KLM postulates. All of these will now be referred to as *LM-rational.* Formally, LM-rationality is defined as follows [26]:

Definition 5.9. *Any defeasible entailment relation* $\approx$ *satisfying the following KLM postulates:*

1. *(LLE) Left logical equivalence:* $\frac{\mathcal{K} \approx \alpha \leftrightarrow \beta,\ \mathcal{K} \approx \alpha \mid\!\sim \gamma}{\mathcal{K} \approx \beta \mid\!\sim \gamma}$

2. *(RW) Right weakening:* $\frac{\mathcal{K} \approx \alpha \rightarrow \beta,\ \mathcal{K} \approx \gamma \mid\!\sim \alpha}{\mathcal{K} \approx \gamma \mid\!\sim \beta}$

3. *(Ref) Reflexivity:* $\mathcal{K} \mid\!\approx \alpha \mid\!\sim \alpha$

4. *And:* $\dfrac{\mathcal{K} \mid\!\approx \alpha \mid\!\sim \beta, \mathcal{K} \mid\!\approx \alpha \mid\!\sim \gamma}{\mathcal{K} \mid\!\approx \alpha \mid\!\sim \beta \wedge \gamma}$

5. *Or:* $\dfrac{\mathcal{K} \mid\!\approx \alpha \mid\!\sim \gamma, \mathcal{K} \mid\!\approx \beta \mid\!\sim \gamma}{\mathcal{K} \mid\!\approx \alpha \vee \beta \mid\!\sim \gamma}$

6. *(CM) Cautious Monotonicity:* $\dfrac{\mathcal{K} \mid\!\approx \alpha \mid\!\sim \gamma, \mathcal{K} \mid\!\approx \alpha \mid\!\sim \beta}{\mathcal{K} \mid\!\approx \alpha \wedge \beta \mid\!\sim \gamma}$

7. *(RM) Rational Monotonicity:* $\dfrac{\mathcal{K} \mid\!\approx \alpha \mid\!\sim \gamma, \mathcal{K} \not\mid\!\approx \alpha \not\mid\!\sim \beta}{\mathcal{K} \mid\!\approx \alpha \wedge \beta \mid\!\sim \gamma}$

is referred to as LM-rational.

This next result follows directly from the representation theorem for ranked interpretations:

Corollary 5.2.1. *A defeasible entailment relation is LM-rational if and only if it is defined by a ranked interpretation.*

Lehmann and Magidor [49] showed that LM-rationality is a necessary condition for any defeasible entailment relation to satisfy. The contention here is that it is not sufficient. By way of example, consider the Flying Dutchman knowledge base $\mathcal{K} \coloneqq \{\mathtt{boat} \mid\!\sim \mathtt{floats}, \mathtt{leaky} \mid\!\sim \mathtt{boat}, \mathtt{leaky} \mid\!\sim \neg\mathtt{floats}, \mathtt{FlyingDutchman} \mid\!\sim \mathtt{boat}, \mathtt{FlyingDutchman} \mid\!\sim \mathtt{leaky}\}$. Then the following ranked interpretation $\mathcal{R}^{\mathcal{K}}$ can be constructed, where meta-atoms are used to represent the atoms in $\mathcal{K}$, such that `b` is `boat`, `f` is `floats`, `l` is `leaky`, and `d` is `FlyingDutchman`:

5	$\overline{\mathtt{b}}\mathtt{lf}\overline{\mathtt{d}}$	$\overline{\mathtt{b}}\mathtt{l}\overline{\mathtt{f}}\mathtt{d}$	$\mathtt{b}\overline{\mathtt{lfd}}$	$\overline{\mathtt{b}}\overline{\mathtt{lf}}\mathtt{d}$	$\overline{\mathtt{bl}}\mathtt{f}\overline{\mathtt{d}}$	$\overline{\mathtt{blf}}\mathtt{d}$
4	$\overline{\mathtt{b}}\mathtt{lfd}$	$\mathtt{b}\overline{\mathtt{l}}\mathtt{fd}$	$\mathtt{blf}\overline{\mathtt{d}}$	$\mathtt{b}\overline{\mathtt{lf}}\mathtt{d}$	$\overline{\mathtt{bl}}\mathtt{fd}$	$\mathtt{bl}\overline{\mathtt{fd}}$
3	$\overline{\mathtt{blfd}}$					
2	$\mathtt{blfd}$					
1	$\mathtt{bl}\overline{\mathtt{f}}\mathtt{d}$					
0	$\mathtt{b}\overline{\mathtt{l}}\mathtt{f}\overline{\mathtt{d}}$					

$\mathcal{R}^{\mathcal{K}}$ represents an LM-rational entailment relation, by the above representation theorem. Then this entailment relation, $\mid\!\approx$, is such that $\mathcal{K} \mid\!\approx \alpha \mid\!\sim \beta$ if and only if $\mathcal{R}^{\mathcal{K}} \Vdash \alpha \mid\!\sim \beta$.

First, note that $\mathcal{R}^{\mathcal{K}}$ is indeed a ranked model of $\mathcal{K}$, that $\mathcal{R}^{\mathcal{K}} \Vdash \mathcal{K}$. Then what inferences are entailed by the knowledge base? Note that $\mathcal{R}^{\mathcal{K}} \Vdash \neg\texttt{FlyingDutchman} \mid\!\sim \texttt{boat}\wedge\texttt{floats}$, that anything that is *not* a Flying Dutchman, is typically a floating boat. Additionally, note that $\mathcal{R}^{\mathcal{K}} \Vdash \texttt{floats} \mid\!\sim \texttt{boat}$, that things that float are typically boats.

The ranked model $\mathcal{R}^{\mathcal{K}}$ represents a pattern of reasoning that is, while perhaps valid from a specific, closed-world, point of view, very flawed. There is no reason for either of the above inferences to follow from the ranked model, especially in light of a defeasible knowledge base which generally represents a common sense view of the world and not any kind of totality of information.

In conclusion, there is a need of some kind of extension of LM-rationality, in order to isolate those ranked models representing a pattern of reasoning that is desirable.

5.4.1 Basic Defeasible Entailment

Since LM-rationality is not enough, the first pass at trying to throw out nonsensical entailment relations is to add some basic properties that constrain what should always be entailed by a defeasible knowledge base. The following three properties all state what a defeasible entailment relation, $\approx$, should *at least* infer:

- Inclusion: for every $\alpha \mid\!\sim \beta \in \mathcal{K}$, $\mathcal{K} \models\!\!\!\approx \alpha \mid\!\sim \beta$
- Classic Preservation: $\mathcal{K} \models\!\!\!\approx \alpha \mid\!\sim \bot$ if and only if $\mathcal{K} \models\!\!\!\approx_P \alpha \mid\!\sim \bot$
- Classic Consistency: $\mathcal{K} \models\!\!\!\approx \top \mid\!\sim \bot$ if and only if $\mathcal{K} \models\!\!\!\approx_P \top \mid\!\sim \bot$

Inclusion simply requires that all defeasible implications in $\mathcal{K}$ are also defeasibly entailed by $\mathcal{K}$, and can be thought of as a defeasible analogue to the Tarskian property of inclusion for classical consequence. Classic preservation states that those defeasible implications that correspond to classical sentences should be exactly the same as those that are preferentially entailed in $\mathcal{K}$. Classic consistency is a direct corollary of classic preservation, by stating that a knowledge base can only be satisfiable if it is satisfiable with respect to preferential entailment.

These properties lead us to the first proper class of defeasible entailment relations:

Definition 5.10. *Any defeasible entailment relation satisfying LM-rationality, Inclusion, and Classic Preservation is referred to as a basic defeasible entailment relation.*

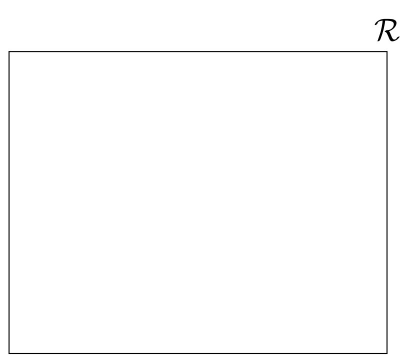
There are a number of derived properties of basic defeasible entailment. An important one is that of $\mathcal{K}$-faithfulness:

Definition 5.11. *Given the ranked interpretation $\mathcal{R}_R^{\mathcal{K}}$ representing the ranked interpretation corresponding to ranked entailment, any ranked model $\mathcal{R}$ of a defeasible knowledge base $\mathcal{K}$ is $\mathcal{K}$-faithful if the set of valuations in $\mathcal{R}$ with non-infinite rank is the same as the set of possible valuations of $\mathcal{K}$. That is, for every $u \in \mathcal{U}$ then $\mathcal{R}(u) \neq \infty$ if and only if $\mathcal{R}_R^{\mathcal{K}}(u) \neq \infty$.*

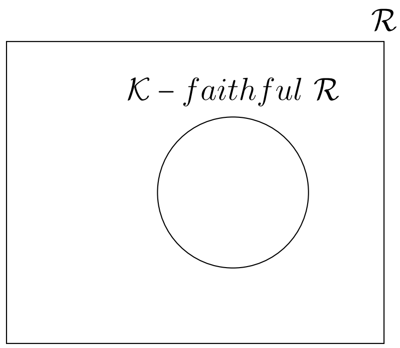
$\mathcal{K}$-faithfulness essentially defines the semantics for basic defeasible entailment: the set of ranked models that are $\mathcal{K}$-faithful define the set of basic defeasible entailment relations.

Recall that the rational closure of a knowledge base $\mathcal{K}$ is defined by the minimal ranked interpretation satisfying $\mathcal{K}$, $\mathcal{R}_{RC}^{\mathcal{K}}$. This ranked interpretation, $\mathcal{R}_{RC}^{\mathcal{K}}$, is $\mathcal{K}$-faithful, and therefore is a basic defeasible entailment relation.

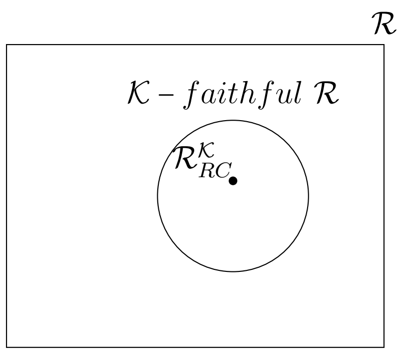
As a visualization of this class of entailment relations, let the following rectangle represent the set of all ranked interpretations, that are therefore LM-rational, given some defeasible knowledge base:



Then, the set of basic defeasible entailments is the subset of these ranked interpretations that are $\mathcal{K}$-faithful:



So, basic defeasible entailment is defined as a subset of the ranked interpretations of a knowledge base, of which the ranked interpretation corresponding to the rational closure of a knowledge base is an element:



Having therefore defined the semantics for basic defeasible entailment, paralleling the rational closure definition, it is possible to define basic defeasible entailment in terms of ranks on formulas, by generalising the notion of base rank, br [26]:

Definition 5.12. *Let $r : \mathcal{L} \mapsto \mathcal{N} \cup \{\infty\}$ be a rank function with $r(\top) = 0$, satisfying the following convexity property: for every $i \in \mathcal{N}$, if $r(\alpha) = i$ then, for every j such that $0 \leq j < i$, there exists $\beta \in \mathcal{L}$ such that $r(\beta) = j$. r is entailment preserving if $\alpha \vDash \beta$ implies $r(\alpha) \geq r(\beta)$. r is $\mathcal{K}$-faithful if:*

- *it is entailment preserving;*
- *$r(\alpha) < r(\alpha \wedge \neg\beta)$ or $r(\alpha) = \infty$, for every $\alpha \mid\!\sim \beta \in \mathcal{K}$, and*
- *$r(\alpha) = \infty$ if and only if $\mathcal{K} \mid\!\approx_{\mathcal{R}} \alpha \mid\!\sim \bot$*

Note, that base rank, $br_{\mathcal{K}}(.)$ is $\mathcal{K}$-faithful. Just as the base rank function can define an entailment relation, so too can any ranking function [26]:

Definition 5.13. *A rank function* $r : \mathcal{L} \mapsto \mathcal{N} \cup \{\infty\}$ *generates a defeasible entailment relation,* $\approx\!\!\!\!|$*, whenever* $\mathcal{K} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta$ *if* $r(\alpha) < r(\alpha \wedge \neg\beta)$ *or* $r(\alpha) = \infty$*.*

Essentially, whereas *br* is a ranking function over formulas that defines the rational closure of a knowledge base, the ranking function r is a more general concept that defines any defeasible entailment relation.

A ranking function, r, is considered to be entailment preserving if its ranking respects the classical logical consequence relationships between the formulas themselves: if $\beta \in \mathcal{L}$ is a logical consequence of $\alpha \in \mathcal{L}$, then β should not be of higher rank than α. Furthermore, it is $\mathcal{K}$-faithful if, in addition to being entailment preserving, for any $\alpha \in \mathcal{L}$ such that $\neg\alpha$ is true in every valuation in every ranked model of $\mathcal{K}$, then $r(\alpha) = \infty$, and if for every defeasible implication $\alpha \mathrel{|\!\sim} \beta \in \mathcal{K}$ the rank assigned to α is strictly lower than the rank of α without β or otherwise the rank of α is infinite. The last criterion can be difficult to understand, but it essentially is a defeasible form of entailment preservation, that forces the ranks to respect the defeasible consequences.

The meaning of the ranks assigned to the formulas corresponds to the meaning of the base rank of a formula. The lower the rank, the more specific the underlying statement it represents. However, whereas the base rank of a formula is a result, the more general rank of a formula from some r is more declarative, and therefore, much like with a ranked interpretation, the ranks a ranking function assigns to formulas in a language encodes a pattern of reasoning, or a particular reading of the knowledge, that directly results in an entailment relation.

Having defined a ranking function, algorithms to compute a basic defeasible entailment can be presented. The first procedure, `Rank` performs the same role as the previous algorithm for rational closure, `BaseRank`, but it is generalized to basic defeasible entailment, and outputs a sequence of sentences, as opposed to how `BaseRank` outputs a sequence of sets of sentences. However, it does require the specific ranking function to be provided, which would correspond to the exact entailment relation to be computed. The algorithm `DefeasibleEntailment` answers whether a query is entailed according to some entailment relation, specified by the `Rank` algorithm. Again, `DefeasibleEntailment` is a generalized version of the `RationalClosure` algorithm. Running `DefeasibleEntailment` with `BaseRank` as an input would result in the rational closure as the output.

Algorithm 3 DefeasibleEntailment

Input: A knowledge base $\mathcal{K}$, a $\mathcal{K}$-faithful rank function r, and a defeasible implication $\alpha \mathrel{|\!\sim} \beta$
Output: **true**, if $\mathcal{K} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta$, and **false** otherwise
$(R_0, ..., R_{n-1}, R_\infty, n) := \text{Rank}(\mathcal{K}, r)$;
$i := 0$
$R := \bigcup_{i=0}^{j<n} \{R_j\}$;
while $\{R_\infty\} \cup R \vDash \neg\alpha$ and $R \neq \emptyset$ **do**
    $R := R \setminus \{R_i\}$;
    $i := i + 1$;
end while
return $\{R_\infty\} \cup R \vDash \alpha \to \beta$;

For the following `Rank` function, let $[\alpha] \coloneqq \{\beta \mid \alpha \equiv \beta, \beta \in \mathcal{L}\}$, that is all formulas logically equivalent to α.

Algorithm 4 Rank

Input: A knowledge base $\mathcal{K}$ and a $\mathcal{K}$-faithful rank function r
Output: An ordered tuple $(R_0, ..., R_{n-1}, R_\infty, n)$
$R_\infty := \neg\Big(\bigvee_{r([\alpha])=\infty} [\alpha]\Big)$;
$n := \max\{i \in \mathbb{N} \mid \exists\alpha \in \mathcal{L} \text{ s.t. } r(\alpha) = i\}$;
if $n = 0$ **then**
    $R_0 := \top$;
    $n := 1$;
else
    for $i := 0$ to $n - 1$ **do**
        $R_i \equiv_{def} \neg\Big(\bigvee_{r([\alpha])=i+1} [\alpha]\Big)$
    end for
end if
return $(R_0, ..., R_{n-1}, R_\infty, n)$

`Rank` produces a tuple of formulas logically equivalent to the material counterparts to the defeasible implications in $\mathcal{K}$, received as input, ordered according to a rank function r, also received as input. The tuple is then used by `DefeasibleEntailment` to determine whether or not $\mathcal{K}$ entails a query taken as input. It does so in the same manner as `RationalClosure`, by removing statements until finding a logically equivalent ranked subset of $\overrightarrow{\mathcal{K}}$ that classically entails the material counterpart of the query. Hence, the output of the `Rank` function may not necessarily look like $\overrightarrow{\mathcal{K}}$, but will be logically equivalent [26]:

Lemma 5.2.1. *Let $(R_0, ..., R_{n-1}, R_\infty, n)$ be the output from the* `Rank` *algorithm, given a defeasible knowledge base $\mathcal{K}$ and a $\mathcal{K}$-faithful rank function $r^{\mathcal{K}}$. Then $\{R_\infty\} \cup \bigcup_{i=0}^{j<n} \{R_j\} \equiv \overrightarrow{\mathcal{K}}$.*

This next theorem is a type of representation theorem for basic defeasible entailment, tying the semantics, the rank function, and the algorithm together.

Theorem 5.3. *The following statements are equivalent:*

1. *$\approx$ is a basic defeasible entailment relation*
2. *there is a $\mathcal{K}$-faithful ranked model $\mathcal{R}$ and a $\mathcal{K}$-faithful rank function r such that:*
   (a) *$r(\alpha) = \min\{i \mid$ there is a $v \in \hat{\alpha}$ such that $\mathcal{R}(v) = i\}$;*
   (b) *$\approx$ can be generated from $\mathcal{R}$;*
   (c) *$\approx$ can be generated from r;*
   (d) *$\approx$ can be computed by* `DefeasibleEntailment`*, given $\mathcal{K}$ and r as input.*

The above theorem ties together the semantics behind basic defeasible entailment, the definition of basic defeasible entailment using ranks on propositions, and the algorithmic definition. A corollary of this theorem, using points 1 and 2b, is that basic defeasible entailment satisfies the following property:

Corollary 5.3.1.

Any basic defeasible entailment relation $\approx$, satisfies the following property, called rank extension: if $\mathcal{K} \approx_R \alpha \mid\sim \beta$, then $\mathcal{K} \approx \alpha \mid\sim \beta$

This is shown by noting that if $\mathcal{K} \approx_R \alpha \mid\sim \beta$, then the defeasible implication $\alpha \mid\sim \beta$ is satisfied by every ranked model of $\mathcal{K}$, including the ranked model that defines $\mathrel{|\!\approx}$. Rank extension defines the monotonic core of $\mathrel{|\!\approx}$, and requires it to extend preferential entailment.

To compare this algorithm to the `RationalClosure` algorithm, consider the knowledge base: $\mathcal{K} \coloneqq \{\texttt{boat} \mathrel{|\!\sim} \texttt{floats}, \neg(\texttt{leaky} \rightarrow \texttt{boat}) \mathrel{|\!\sim} \bot, \texttt{leaky} \mathrel{|\!\sim} \neg\texttt{floats}, \texttt{FlyingDutchman} \mathrel{|\!\sim} \texttt{leaky}\}$, along with the same query as before: $\texttt{FlyingDutchman} \mathrel{|\!\sim} \neg\texttt{floats}$. Then, the material counterpart to $\mathcal{K}$ is $\overrightarrow{\mathcal{K}} \coloneqq \{\texttt{boat} \rightarrow \texttt{floats}, \neg(\texttt{leaky} \rightarrow \texttt{boat}) \rightarrow \bot, \texttt{leaky} \rightarrow \neg\texttt{floats}, \texttt{FlyingDutchman} \rightarrow \texttt{leaky}\}$, and the material counterpart to the query is $\texttt{FlyingDutchman} \rightarrow \neg\texttt{floats}$. Then, consider the following $\mathcal{K}$-faithful ranked interpretation of $\mathcal{K}$, $\mathcal{R}^{\mathcal{K}}$:

∞	$\overline{\texttt{b}}\texttt{l}\overline{\texttt{fd}}$ $\overline{\texttt{b}}\texttt{lfd}$ $\overline{\texttt{b}}\texttt{l}\overline{\texttt{f}}\texttt{d}$ $\overline{\texttt{b}}\texttt{lf}\overline{\texttt{d}}$
3	$\overline{\texttt{bl}}\texttt{fd}$ $\texttt{blfd}$ $\texttt{b}\overline{\texttt{lf}}\texttt{d}$
2	$\texttt{b}\overline{\texttt{l}}\texttt{fd}$ $\texttt{blf}\overline{\texttt{d}}$ $\overline{\texttt{blf}}\texttt{d}$
1	$\texttt{bl}\overline{\texttt{fd}}$ $\texttt{bl}\overline{\texttt{f}}\texttt{d}$ $\texttt{b}\overline{\texttt{lfd}}$
0	$\overline{\texttt{blfd}}$ $\texttt{b}\overline{\texttt{l}}\texttt{f}\overline{\texttt{d}}$ $\overline{\texttt{bl}}\texttt{f}\overline{\texttt{d}}$

$\mathcal{R}^{\mathcal{K}}$ then informs a corresponding $\mathcal{K}$-faithful rank function $r^{\mathcal{K}}$, by theorem 5.3, such that:

- $r^{\mathcal{K}}(\texttt{boat} \wedge \texttt{leaky}) = r^{\mathcal{K}}(\texttt{boat} \wedge \neg\texttt{floats}) = 1$
- $r^{\mathcal{K}}(\texttt{FlyingDutchman} \wedge \neg\texttt{leaky}) = r^{\mathcal{K}}(\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{floats}) = 2$
- $r^{\mathcal{K}}((\texttt{boat} \wedge \texttt{FlyingDutchman}) \wedge (\texttt{leaky} \leftrightarrow \texttt{floats})) = r^{\mathcal{K}}((\texttt{floats} \wedge \texttt{FlyingDutchman}) \wedge \neg\texttt{boat}) = 3$
- $r^{\mathcal{K}}(\neg(\texttt{leaky} \rightarrow \texttt{boat})) = \infty$

Therefore, given $\mathcal{K}$ and $r^{\mathcal{K}}$ as input, `Rank` will produce the following Rs:

- $R_0 \equiv \neg((\texttt{boat} \wedge \texttt{leaky}) \vee (\texttt{boat} \wedge \neg\texttt{floats})) \equiv \neg(\texttt{boat} \wedge \texttt{leaky}) \wedge \neg(\texttt{boat} \wedge \neg\texttt{floats})$
- $R_1 \equiv \neg((\texttt{FlyingDutchman} \wedge \neg\texttt{leaky}) \vee \texttt{boat} \wedge \texttt{leaky} \wedge \texttt{floats}) \equiv (\neg\texttt{FlyingDutchman} \vee \texttt{leaky}) \wedge \neg(\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{floats})$
- $R_2 \equiv \neg(((\texttt{boat} \wedge \texttt{FlyingDutchman}) \wedge (\texttt{leaky} \leftrightarrow \texttt{floats})) \vee ((\texttt{floats} \wedge \texttt{FlyingDutchman}) \wedge \neg\texttt{boat})) \equiv (\neg(\texttt{boat} \wedge \texttt{FlyingDutchman}) \vee \neg(\texttt{leaky} \leftrightarrow \texttt{floats})) \wedge (\neg(\texttt{floats} \wedge \texttt{FlyingDutchman}) \vee \texttt{boat})$

- $R_\infty \equiv \neg(\neg(\texttt{leaky} \rightarrow \texttt{boat})) \equiv \texttt{leaky} \rightarrow \texttt{boat}$

Additionally, `Rank` will output $n = 3$. Then, the `DefeasibleEntailment` algorithm takes as input $\mathcal{K}$, the query $\texttt{FlyingDutchman} \mid\!\sim \neg\texttt{floats}$, and $r^{\mathcal{K}}$, and uses the `Rank` function to obtain $(R_0, R_1, R_2, R_\infty, n)$. Then, similarly to `RationalClosure`, `DefeasibleEntailment` finds the first k, such that $\{R_\infty\} \cup \bigcup_{i=k}^{j<n} \{R_i\} \models \texttt{FlyingDutchman} \rightarrow \neg\texttt{floats}$. First, the algorithm will set $R = R_0 \cup R_1 \cup R_2$. A visualization of $R \cup R_\infty$ is as follows, with `FlyingDutchman` abbreviated as `FD`:

∞	$\texttt{leaky} \rightarrow \texttt{boat}$
2	$(\neg(\texttt{boat} \wedge \texttt{FD}) \vee \neg(\texttt{leaky} \leftrightarrow \texttt{floats})) \wedge (\neg(\texttt{floats} \wedge \texttt{FD}) \vee \texttt{boat})$
1	$(\neg\texttt{FD} \vee \texttt{leaky}) \wedge \neg(\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{floats})$
0	$\neg(\texttt{boat} \wedge \texttt{leaky}) \wedge \neg(\texttt{boat} \wedge \neg\texttt{floats})$

First, `DefeasibleEntailment` checks if $R \cup R_\infty \models \neg\texttt{FlyingDutchman}$. There are only three valuations that satisfy $R \cup R_\infty$, all of which are models of $\neg\texttt{FlyingDutchman}$. Since R is not empty, the algorithm will then remove R_0 from R, leaving the following $R \cup R_\infty$:

∞	$\texttt{leaky} \rightarrow \texttt{boat}$
2	$(\neg(\texttt{boat} \wedge \texttt{FD}) \vee \neg(\texttt{leaky} \leftrightarrow \texttt{floats})) \wedge (\neg(\texttt{floats} \wedge \texttt{FD}) \vee \texttt{boat})$
1	$(\neg\texttt{FD} \vee \texttt{leaky}) \wedge \neg(\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{floats})$

Again, the algorithm will check whether $R \cup R_\infty \models \neg\texttt{FlyingDutchman}$. This time, there is a model of $R \cup R_\infty$, $\texttt{bl}\overline{\texttt{f}}\texttt{d}$, that does not satisfy the formula $\neg\texttt{FlyingDutchman}$, and therefore $R \cup R_\infty \not\models \neg\texttt{FlyingDutchman}$. Then, `DefeasibleEntailment` checks if $R \cup R_\infty \models \texttt{FlyingDutchman} \rightarrow \neg\texttt{floats}$, which is the case as every model of $R_1 \cup R_2 \cup R_\infty$ is a model of $\texttt{FlyingDutchman} \rightarrow \neg\texttt{floats}$, and therefore `DefeasibleEntailment` returns **true**.

Now, having defined the class of basic defeasible entailment, does it satisfy the goal laid out at the beginning of this section by removing all nonsensical defeasible entailments? It appears that while it is a good effort, it is still too permissive. To illustrate why, an example [26]:

2	$\texttt{pbf}$
1	$\overline{\texttt{p}}\texttt{b}\overline{\texttt{f}}$ $\texttt{pb}\overline{\texttt{f}}$
0	$\overline{\texttt{pbf}}$ $\overline{\texttt{pb}}\texttt{f}$ $\overline{\texttt{p}}\texttt{bf}$

2	$\mathtt{pbf}$
1	$\overline{\mathtt{pbf}}$ $\overline{\mathtt{p}}\mathtt{b}\overline{\mathtt{f}}$ $\mathtt{pb}\overline{\mathtt{f}}$
0	$\overline{\mathtt{pb}}\mathtt{f}$ $\overline{\mathtt{p}}\mathtt{bf}$

Above are two ranked interpretations of the knowledge base $\mathcal{K} \coloneqq \{p \rightarrow b, b \mathrel{|\!\sim} f, p \mathrel{|\!\sim} \neg f\}$ [26]. One of which, $\mathcal{R}^{\mathcal{K}}_{RC}$, is the minimal ranked model, corresponding to the rational closure, and the other, $\mathcal{R}^{\mathcal{K}}$ is a $\mathcal{K}$-faithful ranked model, which corresponds to a basic defeasible entailment relation. Now, $\mathcal{R}^{\mathcal{K}}_{RC} \Vdash \neg p \wedge \neg f \mathrel{|\!\sim} \neg\beta$, which means that $\mathcal{K} \mathrel{|\!\approx}_{RC} \neg p \wedge \neg f \mathrel{|\!\sim} \neg b$, i.e., that objects that are not ps and not fs are typically not bs either is in the rational closure of $\mathcal{K}$. However, note that the $\mathcal{K}$-faithful ranked model does not agree. $\mathcal{R}^{\mathcal{K}} \nVdash \neg p \wedge \neg f \mathrel{|\!\sim} \neg b$, and therefore the basic defeasible entailment relation defined by the $\mathcal{K}$-faithful ranked model does not entail this defeasible implication. The takeaway is that basic defeasible entailment does not guarantee that at least the defeasible implications entailed by rational closure will be inferred. The significance of this is related to the significance of rational closure as a nonmonotonic core for defeasible entailment [26]. Rational closure is defined by the minimal ranked interpretation of a knowledge base, and in that sense represents the minimum that should be entailed by a nonmonotonic entailment. Any entailment relation that entails at least as much as the rational closure should be a reasonable entailment relation, but basic defeasible entailment does not guarantee that, which implies that a subset of basic defeasible entailment will entail not much more than ranked entailment. This implies that a stronger class of defeasible entailment relations should be defined, one that enforces rational closure as a nonmonotonic core.

5.4.2 Rational Defeasible Entailment

Given the argument for rational closure as a nonmonotonic core of defeasible entailment, it is natural to formulate it as a core property for the next class of defeasible entailment:

Definition 5.14.

Rational Closure Extension (RC Extension): given a defeasible knowledge base $\mathcal{K}$ and a defeasible entailment relation $\mathrel{|\!\approx}$, it holds that $\mathrel{|\!\approx}$ satisfies rational closure extension if it is the case that if $\mathcal{K} \mathrel{|\!\approx}_{RC} \alpha \mathrel{|\!\sim} \beta$ then $\mathcal{K} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta$.

RC extension simply requires that a defeasible entailment relation extends the rational closure of $\mathcal{K}$. If a basic defeasible entailment relation satisfies

RC Extension, then it is a *rational* defeasible entailment relation. Rational defeasible entailment relations can be characterised semantically as a subset of $\mathcal{K}$-faithful ranked interpretations satsifying the following property:

Definition 5.15.

Rank Preservation: A ranked interpretation $\mathcal{R}$ is rank preserving if for all $v, u \in \mathcal{U}$ if $\mathcal{R}_{\mathcal{K}}^{RC}(v) < \mathcal{R}_{\mathcal{K}}^{RC}(u)$ then $\mathcal{R}_{\mathcal{K}}(v) < \mathcal{R}_{\mathcal{K}}(u)$

Rank preservation is an interesting property, and, informally, states that a rank preserving ranked model is one that respects the relative rankings between each pair of valuations defined in the minimal ranked interpretation $\mathcal{R}_{RC}^{\mathcal{K}}$.

Naturally, rational defeasible entailment can then also be characterised in terms of the rank function r if it satisfies the corresponding property of being *base rank preservation*:

Definition 5.16. *A rank function r is said to be base rank preserving if for all $\alpha, \beta \in \mathcal{L}$, $br_{\mathcal{K}}(\alpha) < br_{\mathcal{K}}(\beta)$, then $r(\alpha) < r(\beta)$.*

Base rank preserving rank functions assign ranks to formulas that respect the relative rankings between formulas assigned by the base rank function, analogously to *rank preservation* described previously for the semantics.

To illustrate *rank preservation*, consider the following two ranked models, with valuations of infinite rank omitted for brevity, of the knowledge base $\mathcal{K} := \{\neg(p \to b) \mid\!\sim \bot,\ b \mid\!\sim f,\ p \mid\!\sim \neg f\}$:

2	$\mathtt{pbf}$
1	$\overline{\mathtt{p}}\mathtt{b}\overline{\mathtt{f}}$ $\mathtt{pb}\overline{\mathtt{f}}$
0	$\overline{\mathtt{pbf}}$ $\overline{\mathtt{pb}}\mathtt{f}$ $\overline{\mathtt{p}}\mathtt{bf}$

3	$\mathtt{pbf}$
2	$\overline{\mathtt{p}}\mathtt{b}\overline{\mathtt{f}}$ $\mathtt{pb}\overline{\mathtt{f}}$
1	$\overline{\mathtt{pb}}\mathtt{f}$ $\overline{\mathtt{p}}\mathtt{bf}$
0	$\overline{\mathtt{pbf}}$

The top ranked model is the minimal ranked model of $\mathcal{K}$, $\mathcal{R}_{RC}^{\mathcal{K}}$, and the bottom ranked model is a constructed ranked model of $\mathcal{K}$, $\mathcal{R}^{\mathcal{K}}$, defining a rational defeasible entailment relation. Note that for every $u, v \in \mathcal{U}$, wherever

$\mathcal{R}^{\mathcal{K}}_{RC}(u) < \mathcal{R}^{\mathcal{K}}_{RC}(v)$ then also $\mathcal{R}^{\mathcal{K}}(u) < \mathcal{R}^{\mathcal{K}}(v)$. Therefore, $\mathcal{R}^{\mathcal{K}}$ is a *rank-preserving* ranked model of $\mathcal{K}$. This leads to the primary theorem defining rational defeasible entailment:

Theorem 5.4. *The following statements are equivalent:*

1. $\approx\!\!|$ *is a rational defeasible entailment relation*
2. *there is a rank preserving* $\mathcal{K}$*-faithful ranked interpretation* $\mathcal{R}$ *and a base rank preserving* $\mathcal{K}$*-faithful rank function* r *such that:*
   (a) $r(\alpha) = \min\{i \mid v \in [\![\alpha]\!] \text{ and } \mathcal{R}(v) = i\}$;
   (b) $\approx\!\!|$ *can be generated from* $\mathcal{R}$;
   (c) $\approx\!\!|$ *can be generated from* r;
   (d) $\approx\!\!|$ *can be computed from* `DefeasibleEntailment`*, given* $\mathcal{K}$ *and* r *as input.*

The above theorem is completely analogous to the theorem defining basic defeasible entailment, the only changes are the new properties defined with which r and $\mathcal{R}$ should comply. Of note, is that the previously defined algorithm `DefeasibleEntailment`, can of course also compute any rational defeasible entailment relation given an appropriate, $\mathcal{K}$-faithful base rank preserving rank function.

5.5 Other Nonmonotonic Entailment Relations

5.5.1 Lexicographic Closure

Lehmann [48] proposed an entailment relation using the framework defined by Lehmann and Magidor [49] to formalize the pattern of default reasoning as described by Reiter [61]. First, Lehmann described informal, intuitive properties that underpin default reasoning, and that should therefore guide the definition of lexicographic closure. They are:

1. the presumption of typicality
2. the presumption of independence
3. priority to typicality

4. respect for specificity

The presumption of typicality can be thought of as a stronger assumption of monotonicity in the absence of any information to the contrary. Rational monotonicity states that it is possible to accept new information without changing a conclusion on the condition that the new information does not contradict with any existing information. However, it does not unambiguously inform what should be derived, due to its nature as a negative inference rule. Given a knowledge base $\mathcal{K} \coloneqq \{\texttt{boat} \mid\!\sim \texttt{floats}\}$ and an LM-rational entailment relation $\approx$, both $\mathcal{K} \approx \texttt{boat} \wedge \texttt{sailors} \mid\!\sim \texttt{floats}$ or $\mathcal{K} \approx \texttt{boat} \mid\!\sim \neg\texttt{sailors}$ could be acceptable inferences using rational monotonicity. The presumption of typicality states that the former should be preferred, as there is no strong justification to accept the latter instead [48].

The presumption of independence strengthens the tendency towards monotonicity established by the presumption of typicality. The presumption of independence broadly says to presume typicality for every consequent unless there is a reason for the contrary. Following from the previous example, if there is a $\mathcal{K} \coloneqq \{\texttt{boat} \mid\!\sim \texttt{floats}, \texttt{boat} \mid\!\sim \neg\texttt{sailors}\}$ then the presumption of typicality cannot be used to support $\mathcal{K} \approx \texttt{boat} \wedge \texttt{sailors} \mid\!\sim \texttt{floats}$ since there is already explicitly $\texttt{boat} \mid\!\sim \neg\texttt{sailors}$ in the knowledge base. The presumption of independence states that since the relationship between `sailors` and `floats` is not known, it should be assumed that `sailors` has no relation to `floats`, and therefore $\texttt{boat} \wedge \texttt{sailors} \mid\!\sim \texttt{floats}$ should be accepted, despite the fact that $\neg\texttt{sailors}$ is known to be a typical consequence of `boat`.

Priority to typicality resolves conflict between the presumption of typicality and presumption of independence. Given two inferences, one derived from the presumption of typicality and the other derived from presumption of independence, priority to typicality states that the former should be preferred. Consider the knowledge base $\mathcal{K} \coloneqq \{\texttt{boat} \mid\!\sim \texttt{floats}, \texttt{boat} \wedge \texttt{leaky} \mid\!\sim \neg\texttt{floats}\}$ then the presumption of typicality suggests that $\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{sailors} \mid\!\sim \neg\texttt{floats}$ is inferred, and the presumption of independence offers both $\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{sailors} \mid\!\sim \neg\texttt{floats}$ and also $\texttt{boat} \wedge \texttt{leaky} \wedge \texttt{sailors} \mid\!\sim \texttt{floats}$. Since it is not justified to accept both, as deriving both a formula and its negation results in an unsatisfiable knowledge base, priority to typicality concludes that the former should hold.

Lastly, respect for specificity states that given any two clashing inferences, the inference based on the defeasible implication with a more specific antecedent should be preferred. The exact nature of what it means for an inference to be "based on" a defeasible implication is difficult to formalize, and is guided mostly by intuition. Furthermore, there seems to be a close relationship between this principle and the priority to typicality. Following the priority to typicality generally gives us exactly the results one would expect from respect for specificity. Note in the example above for priority to typicality, the accepted inference was derived from the more specific $\texttt{boat} \wedge \texttt{leaky} \mid\sim \neg\texttt{floats}$ over $\texttt{boat} \mid\sim \texttt{floats}$. Intuitively, a leaky boat is more specific than a, presumably, regular boat. Formally, $\texttt{boat} \wedge \texttt{leaky}$ is more information than $\texttt{boat}$, and so clearly is more specific. However, not all cases are so clear. It is possible to have a knowledge base such as $\mathcal{K} \coloneqq \{\texttt{bird} \mid\sim \texttt{fly}, \texttt{penguin} \mid\sim \texttt{bird} \wedge \neg\texttt{fly}, \texttt{tweety} \wedge \texttt{bird} \mid\sim \texttt{fly}\}$, where specificity is not as obvious. Is $\texttt{tweety}$ on the same level of specificity as $\texttt{penguin}$? It is not, it is less specific, but $\texttt{tweety} \wedge \texttt{bird}$ is as specific as $\texttt{penguin}$, which is less obvious by just looking at the formulas.

Another, more informal principle worth mentioning is dubbed *avoidance of junk*. This principle states that the closed set of statements produced by an entailment relation on a knowledge base should be minimal in the set-theoretic sense of the term: that no strict subset should be acceptable. The essential meaning of this principle is that spurious inferences should be avoided.

Recall that the base rank function br is a ranking function defined to describe rational closure. Then for any defeasible knowledge base, the order of the knowledge base is defined as [48]:

Definition 5.17. *Given a defeasible knowledge base $\mathcal{K}$, the order of a knowledge base, k, is the maximum base rank assigned to a formula in the knowledge base, not including ∞, such that $k^{\mathcal{K}} \coloneqq \max\{br(\alpha) \mid \alpha \mid\sim \beta \in \mathcal{K}, br(\alpha) \in \mathcal{N}\}$.*

The aim now is to find a ranked interpretation, $\mathcal{R}_{LC}$, that corresponds to the lexicographic closure. Finding a ranked interpretation is essentially finding a modular ordering, that shall be denoted $\prec^{LC}$, over a set of valuations $V \subseteq \mathcal{U}$. It should be clear that there will be valuations that violate some subset of defeasible implications, $D \subseteq \mathcal{K}$, and so the modular ordering over these valuations will be informed by ordering the subsets of defeasible implications violated. To order these subsets, a measure of "seriousness" needs to be defined, violating a more "serious" set of defeasible implications is less preferable, therefore a valuation that violates a less serious subset is preferable to a valuation that violates a more serious subset of D. Seriousness is defined on the following two metrics [48]:

- the size of the set: violating a subset with a smaller cardinality is less serious than violating that of a larger cardinality.
- the specificity of the elements: it is less serious to violate a less specific statement than a more specific one. Specificity here is informed by the base rank of the statement.

Having defined two metrics to provide the ordering $\prec^{LC}$, the question now is how to compose them into a single ordering? Here is where this entailment relation gets its name: a lexicographic composition, where the second criterion, specificity, is treated as the major discriminator, and then the first criterion as a secondary discriminator when two valuations are equivalent according to the first. The motivation here is that it should be preferable to violate two defeasible statements of low specificity to violating a single one of higher specificity.

Now that the intuition behind such a "seriousness" ordering has been described, what follows is a more formal definition. First, a definition for the seriousness of subsets of defeasible implications. Note that for any set X, then $|X|$ denotes the cardinality of X:

Definition 5.18. *Given a defeasible knowledge base $\mathcal{K}$ of order k, every subset $D \subseteq \mathcal{K}$ has a corresponding $k+1$ tuple of natural numbers, denoted n_D, $\langle n_0, ..., n_k \rangle$, where each is defined as such: $n_0 = |\{\alpha \mid\sim \beta \in D \mid br_{\mathcal{K}}(\alpha) = \infty\}|$, $n_1 = |\{\alpha \mid\sim \beta \in D \mid br_{\mathcal{K}}(\alpha) = k-1\}|$ and for any $i = 1, ..., k$, $n_i = |\{\alpha \mid\sim \beta \in D \mid br_{\mathcal{K}}(\alpha) = k-i\}|$. That is, n_0 is the number of defeasible implications of infinite base rank, or having no rank, in D, and each n_i for $0 < i \leq k$ is the number of defeasible implications of base rank $k-i$ in D.*

This definition gives, for any set of defeasible information D a corresponding tuple n_D, that can be used to compare D to some other set of defeasible implications, B, and allows for a preference order that is defined, as mentioned above, lexicographically. Starting from the highest rank, ∞, to the lowest rank 0, if at any point there is a higher number in n_D than n_B for a particular rank, then $B \prec_S D$. This can be formalized as follows:

Definition 5.19. *Given a defeasible knowledge base $\mathcal{K}$, a seriousness ordering on subsets $D \subseteq \mathcal{K}$ is a modular partial ordering, denoted $\prec_S$, that is a lexicographic ordering over the tuples of ranks for each subset. That is, given two subsets, $D_1, D_2 \subseteq \mathcal{K}$, $D_1 \prec_S D_2$ if and only if $n_{D_1} \prec_S n_{D_2}$ using the natural lexicographic ordering over tuples of natural numbers, e.g. $\langle 1,0,2 \rangle \prec_S \langle 1,1,2 \rangle$. $\prec_S$ is a strict modular partial order over subsets of $\mathcal{K}$.*

The intuition behind the seriousness ordering is that for two subsets $D_1, D_2 \subseteq \mathcal{K}$, if $D_1 \prec_S D_2$ then there is a base rank i for which D_2 has a higher number of formulas, and for all ranks higher than i, including ∞, D_1 and D_2 have the same number of formulas.

This seriousness ordering over subsets of a defeasible knowledge base $\mathcal{K}$ allows for the definition of a *basis* [48] for a formula in $\mathcal{K}$:

Definition 5.20. *Given a defeasible knowledge base $\mathcal{K}$, and a formula α, then a subset $B \subseteq \mathcal{K}$, is referred to as a basis for α if and only if the material counterpart of B, $\overrightarrow{B}$, is consistent with α, i.e., $\overrightarrow{B} \not\models \neg\alpha$, and also B is maximal with respect to the seriousness ordering, i.e., there is no B' such that $B \prec_S B'$ where $\prec_S$ is the seriousness ordering.*

For some formula α, a basis is a subset of the defeasible knowledge base such that α is consistent, i.e., its negation is not entailed, and, if $br(\alpha) = i$, then for any $j \geq i$, it is the case that $\mathcal{E}_j^{\mathcal{K}} \subseteq B$, i.e., that B contains at least all statements in $\mathcal{K}$ of rank equal to or higher than α.

Using this definition of a basis, lexicographic closure can be defined using the following theorem from Lehmann [48]:

Theorem 5.5. *Given a defeasible knowledge base $\mathcal{K}$, and a defeasible implication $\alpha \mid\!\sim \beta$, it is the case that $\alpha \mid\!\sim \beta$ is in the lexicographic closure of $\mathcal{K}$, denoted $\mathcal{K} \mid\!\approx_{LC} \alpha \mid\!\sim \beta$, if and only if for any basis $B \subseteq \mathcal{K}$ of α, $\overrightarrow{B} \cup \{\alpha\} \models \beta$.*

The seriousness ordering defined previously can also be used to construct a ranked interpretation of $\mathcal{K}$ that corresponds to lexicographic closure. This can be formalized in the following definition from Lehmann [48] as well as Casini et al. [26]:

Definition 5.21. *Given a defeasible knowledge base $\mathcal{K}$, and valuations $m, n \in \mathcal{U}$, the preference order $\prec_{LC}$ over $\mathcal{U}$ is defined as: $m \prec_{LC} n$ if and only if $V(m) \prec_S V(n)$ where $V(m) \subseteq \mathcal{K}$ is the set of defeasible implications violated by $m \in \mathcal{U}$. $\prec_{LC}$ is a modular partial order over $\mathcal{U}$, and so defines a ranked interpretation, denoted $\mathcal{R}_{LC}^{\mathcal{K}}$. $\mathcal{R}_{LC}^{\mathcal{K}}$ is $\mathcal{K}$-faithful and rank preserving.*

This ranked interpretation defines an entailment relation that is exactly lexicographic closure:

Definition 5.22. *Given a knowledge base $\mathcal{K}$, a defeasible implication $\alpha \mid\!\sim \beta$, and a ranked interpretation $\mathcal{R}_{LC}^{\mathcal{K}}$ of $\mathcal{K}$, it holds that $\mathcal{K} \mid\!\approx_{LC} \alpha \mid\!\sim \beta$ if and only if $\mathcal{R}_{LC}^{\mathcal{K}} \Vdash \alpha \mid\!\sim \beta$.*

Since lexicographic closure can be defined by a ranked interpretation, it is therefore LM-rational, and was shown to satisfy *RC Extension* by Lehmann [48], by noting that the previously defined seriousness order can also be used to define rational closure.

It also follows from this that there should be a rank function that can be used to generate the lexicographic closure [26]:

Definition 5.23. *The lexicographic rank function, $r^{LC}_{\mathcal{K}}$, with respect to a knowledge base $\mathcal{K}$, is defined as $r^{LC}_{\mathcal{K}}(\alpha) \coloneqq \min\{\mathcal{R}^{\mathcal{K}}_{LC}(v) \mid v \in \hat{\alpha}\}$.*

The rank function r^{LC} is both $\mathcal{K}$-faithful and base rank preserving, which follows from the definition of lexicographic closure, as well as from lexicographic closure satisfying *RC Extension*. Then, it follows that lexicographic closure can be computed with the same algorithms that compute basic defeasible entailments:

Lemma 5.5.1. *The algorithm* `DefeasibleEntailment` *returns* **true** *when given the input $\mathcal{K}$, $r^{LC}_{\mathcal{K}}$, and $\alpha \mid\!\sim \beta$ if and only if $r^{LC}_{\mathcal{K}}(\alpha) < r^{LC}_{\mathcal{K}}(\alpha \wedge \neg\beta)$, or $r^{LC}_{\mathcal{K}}(\alpha) = \infty$.*

These various definitions of lexicographic closure can be tied up as follows:

Theorem 5.6. *Given a defeasible knowledge base $\mathcal{K}$, and a defeasible implication $\alpha \mid\!\sim \beta$, then the following statements are equivalent:*

- $\mathcal{K} \mathrel{|\!\approx}_{LC} \alpha \mid\!\sim \beta$
- $\mathcal{R}^{\mathcal{K}}_{LC} \Vdash \alpha \mid\!\sim \beta$
- $r^{LC}_{\mathcal{K}}(\alpha) < r^{LC}_{\mathcal{K}}(\alpha \wedge \neg\beta)$ *or* $r^{LC}_{\mathcal{K}}(\alpha) = \infty$
- *The algorithm* `DefeasibleEntailment` *returns* **true** *when given the input $\mathcal{K}$, $r^{LC}_{\mathcal{K}}$, and $\alpha \mid\!\sim \beta$.*

For an example, consider the knowledge base $\mathcal{K} \coloneqq \{\texttt{boat} \mid\!\sim \texttt{floats}, \texttt{boat} \mid\!\sim \texttt{sailors}, \neg(\texttt{leaky} \to \texttt{boat}) \mid\!\sim \bot, \texttt{leaky} \mid\!\sim \neg\texttt{floats}\}$. Then the ranked model, with valuations of infinite rank omitted, corresponding to the lexicographic closure, $\mathcal{R}^{\mathcal{K}}_{LC}$, is the following:

5	$\mathtt{blf\overline{s}}$
4	$\mathtt{blfs}$
3	$\mathtt{bl\overline{fs}}$ $\mathtt{b\overline{lfs}}$
2	$\mathtt{bl\overline{f}s}$ $\mathtt{b\overline{lf}s}$
1	$\mathtt{b\overline{l}f\overline{s}}$
0	$\mathtt{\overline{blfs}}$ $\mathtt{\overline{blf}s}$ $\mathtt{\overline{bl}f\overline{s}}$ $\mathtt{\overline{bl}fs}$ $\mathtt{b\overline{l}fs}$

While the ranked model corresponding to the rational closure, $\mathcal{R}^{\mathcal{K}}_{RC}$, is:

2	$\mathtt{blf\overline{s}}$ $\mathtt{blfs}$
1	$\mathtt{bl\overline{f}s}$ $\mathtt{bl\overline{fs}}$ $\mathtt{b\overline{lf}s}$ $\mathtt{b\overline{lfs}}$
0	$\mathtt{b\overline{l}f\overline{s}}$ $\mathtt{\overline{blfs}}$ $\mathtt{\overline{blf}s}$ $\mathtt{\overline{bl}f\overline{s}}$ $\mathtt{\overline{bl}fs}$ $\mathtt{b\overline{l}fs}$

Visually, the difference between the two ranked models is quite clear. $\mathcal{R}^{\mathcal{K}}_{LC}$ is more "stretched out" than $\mathcal{R}^{\mathcal{K}}_{RC}$, as lexicographic closure has a more fine grained preference ordering than rational closure. This fine grained ordering is what allows for a more presumptive reading. Consider the query `leaky` $\mid\sim$ `sailors`. The minimal `leaky` valuations in $\mathcal{R}^{\mathcal{K}}_{RC}$ are $\mathtt{bl\overline{f}s}$ $\mathtt{bl\overline{fs}}$, and so $\mathcal{K} \not\approx_{RC}$ `leaky` $\mid\sim$ `sailors`, however the minimal `leaky` valuation in $\mathcal{R}^{\mathcal{K}}_{LC}$ is $\mathtt{bl\overline{f}s}$, and so it is the case that $\mathcal{K} \approx_{LC}$ `leaky` $\mid\sim$ `sailors`, which illustrates the distinction between rational closure and lexicographic closure, and the effect of a finer grained preference order.

5.5.2 Prototypical and Presumptive Entailment

Rational closure is what is termed a *prototypical* reading of a knowledge base. What is meant by this is that rational closure will infer as few a number of defeasible inferences relative to other ranked interpretations, a product of rational closure corresponding to the minimal ranked interpretation, and so is the most prototypical defeasible entailment relation, representing the most conservative pattern of defeasible reasoning. It can also be seen in comparing the *br* ranking function to other general ranking functions r. *br* is the function that "pushes down" the ranks of the formulas as much as possible, as a result of $\mathcal{R}_{RC}$, which pushes down the ranks of the valuation the most. Prototypical reading is called as such since it describes how $\mid\sim$ is interpreted. Under a prototypical reading, $\alpha \mid\sim \beta$ is read as "typically, given α it is possible to conclude that β is the case". This interpretation informs why rational closure draws as few inferences as possible: if there is some γ that is an abnormal α, rational closure concludes that it is not possible to conclude any $\gamma \mid\!\sim \beta$ for any β on the basis that $\alpha \mid\!\sim \beta$ is derivable from the knowledge base, as the prototypical reading states that it is only from a *typical* α from which β can be inferred.

On the other hand, there is a *presumptive* reading of $\mid\!\sim$, which means that $\alpha \mid\!\sim \beta$ is read as "from α, β can be presumed unless stated otherwise", which is a fundamental difference in the interpretation of $\mid\!\sim$. Lexicographic closure represents a presumptive reading of $\mid\!\sim$, as it is willing to conclude as much as possible from even an atypical α. For example, lexicographic closure will conclude that a penguin has wings, despite being an abnormal, in this case flightless, bird.

A natural question is whether lexicographic closure is the most presumptive entailment. The immediate answer is that it is not. Consider the ranked interpretation above from which lexicographic closure is generated, it is clear that it is not the most granulated ranked interpretation that could be constructed. Consider the ranked interpretation, $\mathcal{R}^{\mathcal{K}}$, below of the same knowledge base, $\mathcal{K} \coloneqq \{\mathtt{boat} \mid\!\sim \mathtt{floats}, \mathtt{boat} \mid\!\sim \mathtt{sailors}, \neg(\mathtt{leaky} \rightarrow \mathtt{boat}) \mid\!\sim \bot, \mathtt{leaky} \mid\!\sim \neg\mathtt{floats}\}$, defined previously:

7	$\mathtt{blf}\overline{\mathtt{s}}$
6	$\mathtt{blfs}$
5	$\mathtt{bl}\overline{\mathtt{fs}}$
4	$\mathtt{b}\overline{\mathtt{lfs}}$
3	$\mathtt{bl}\overline{\mathtt{f}}\mathtt{s}$
2	$\mathtt{b}\overline{\mathtt{lf}}\mathtt{s}$
1	$\mathtt{b}\overline{\mathtt{l}}\mathtt{f}\overline{\mathtt{s}}$
0	$\overline{\mathtt{blfs}}$ $\overline{\mathtt{blf}}\mathtt{s}$ $\overline{\mathtt{bl}}\mathtt{f}\overline{\mathtt{s}}$ $\overline{\mathtt{bl}}\mathtt{fs}$ $\mathtt{b}\overline{\mathtt{l}}\mathtt{fs}$

The above ranked model of $\mathcal{K}$ is a refinement of the ranked model $\mathcal{R}^{\mathcal{K}}_{LC}$, and as such is $\mathcal{K}$-faithful and base rank preserving, and therefore generates a rational defeasible entailment relation, $\approx_{\mathcal{R}}$. It can be shown that $\approx_{\mathcal{R}}$ extends lexicographic closure, i.e., that if $\mathcal{K} \approx_{LC} \alpha \mid\!\sim \beta$ then $\mathcal{K} \approx_{\mathcal{R}} \alpha \mid\!\sim \beta$ by comparing it to $\mathcal{R}^{\mathcal{K}}_{LC}$ in the previous section, and is, in fact, more presumptive than lexicographic closure. For example, note that $\mathcal{K} \approx_{\mathcal{R}} \mathtt{boat} \wedge \neg\mathtt{floats} \wedge \mathtt{sailors} \mid\!\sim \neg\mathtt{leaky}$, that a boat with sailors that does not float is typically not leaky, but $\mathcal{K} \not\approx_{LC} \mathtt{boat} \wedge \neg\mathtt{floats} \wedge \mathtt{sailors} \mid\!\sim \neg\mathtt{leaky}$.

Prototypical and presumptive entailment relations form a spectrum of patterns of reasoning. Rational closure has been shown to be the most prototypical entailment relation and so forms the one end of the spectrum, however it has been demonstrated that lexicographic closure is not the most presumptive. The implication is that the number of even just rational defeasible entailment relations is quite large, and given a defeasible knowledge base there are many ranked models that can be constructed, with each one representing a defeasible entailment relation with a different reading of $\mid\!\sim$, depending on whether it represents information about the way things typically are, or the way things may be presumed to be.

5.5.3 Other Types of Closure

Rational closure and lexicographic closure are the entailment relations defined by KLM [46] and Lehmann [48], respectively, in this framework. However, the framework has been shown to be far broader than that. As previously stated, rational closure is defined as being the *most* prototypical consequence relation, and lexicographic closure a presumptive relation. The natural question is then: what various entailment relations are there, and how do they differ? Rational entailment, as presented previously, is intended as a guideline for defining any entailment relation corresponding to some pattern of reasoning. However, what will now be presented is an example of an entailment constructed differently to what has been described, but still using the KLM framework.

5.5.4 Relevant Closure

An interesting entailment relation is relevant closure, first described by Casini et al.[25]. Consider the problem where a subset of properties should be inherited by an atypical example. Given that birds have wings and fly, and that an ostrich is a bird that does not fly, we would ideally want to inherit that they still have wings, even though they are an atypical bird. Rational closure will not conclude that ostriches have wings, as it is the most prototypical reading of the knowledge. There are a number of other presumptive entailments that are willing to draw such an inference, however relevant closure provides an elegant, different reading that attempts to respect where rational closure is correct to be prototypical, but still add presumptive inferences that are based on a notion of *relevance* attached to each statement.

This puts it somewhere in between rational closure and lexicographic closure on the spectrum of prototypical and presumptive entailments.

Relevant closure is defined algorithmically and operates by identifying a subset of defeasible statements that are relevant to a given query, and applies a version of the `DefeasibleEntailment` algorithm to the knowledge base such that only those relevant statements may be removed. Given a defeasible knowledge base $\mathcal{K}$, and a query $\alpha \mid\sim \beta$, then identify some $R \subset \mathcal{K}$ that is the subset of defeasible statements relevant to the query. R therefore is partitioned by the base rank of the formulas, the same base rank, br, that defines the rational closure, and each rank is removed in the usual way as in `DefeasibleEntailment` until there is some rank of R together with $\mathcal{K} \setminus R$ that entails the query, and if there is no such rank then it is not entailed.

However, relevant closure has some limitations. Firstly, it is not LM-rational, as it does not satisfy *Or*, *Cautious Monotonicity*, and *Rational Monotonicity*. This is not inherently undesirable, but means that the preferential semantics for defeasible entailment presented here does not describe relevant closure.

Relevant closure is emblematic of the utility of KLM-style frameworks. Not every entailment relation need satisfy every postulate, nor does it need to be defined in every available way. Rather, as long as the pattern of reasoning is sound, or useful, then it may be defined semantically, algorithmically, proof theoretically, or some combination of the above using the same formalism.

Chapter 6

Discussion, Conclusion and Future Work

This dissertation has covered classical propositional logic, a class of preferential consequence relations over a propositional language, and then defined a defeasible propositional logic over which a nonmonotonic entailment relation may be defined. Such entailment relations have been demarcated into the various classes of LM-rational, basic defeasible entailment, and rational defeasible entailment dependent on the sets of properties they satisfy. The intuition presented here is that each entailment relation corresponds to some pattern of reasoning that would fall under the broad category of "common sense" reasoning. Not all such common sense reasoning is valid, however, and may well be nonsensical in practice. The framework of rational defeasible entailment is then proposed as isolating a set of entailment relations that are all reasonable, given the correct context. A natural question is then: are there desirable entailment relations that are not rational or perhaps not even a basic defeasible entailment relation? Immediately, relevant closure should be recalled. Relevant closure is not LM-rational, and so is not a basic defeasible entailment relation, and therefore is not a rational defeasible entailment relation. However, the pattern of reasoning it encodes is reasonable, and it does extend the rational closure. What this should illustrate, is that this framework can be treated somewhat flexibly, as relevant closure itself is defined using this framework, but does not fit in the classes defined thus far.

6.1 Nonmonotonicity and defeasibility

Broadly, there are two types of reasoning at play in nonmonotonic reasoning: *defeasible* reasoning, and *ampliative* reasoning [21]. A defeasible inference is simply one that may be later retracted. Defeasibility can then be seen as a meta-level notation over a meta-level or object-level statement: explicitly noting that an inference may be retracted on learning new information, or that this explicit information represents knowledge that holds in the most typical, or preferred, situations, but may not be the case in more atypical situations. This added expressivity is what translates to "common-sense" statements about the way things usually are, rather than classical statements that necessarily hold in all cases.

The other side of this nonmonotonic coin is ampliative reasoning [21]. Ampliativity is the property of inferring more information than another entailment relation given the same knowledge. Compare presumptive entailments, such as lexicographic closure, inferring more statements than the rational closure. The exact relationship between ampliativity and defeasibility can be challenging to define. Intuitively, they seem to be heavily related, and in some way dependent on each other. Ampliativity implies a level of defeasibility: any ampliative inference likely should be defeasible, as its nature implies that it is already inferred on less than concrete grounds, and it is highly likely that there exists additional information that could cause a contradiction. Then, does defeasibility imply ampliativity? This is perhaps a harder question. It is possible to conceive of a defeasible statement that is not particularly ampliative, consider a defeasible inference to a defeasible knowledge base that is logically equivalent to a classical statement, but the real murkiness is the exact definition of an ampliative statement. If ampliativity is defined as being those inferences not in the monotonic core of a knowledge base, which is relatively intuitive, then there do exist defeasible statements that are not ampliative, which suggests that defeasibility does not quite imply ampliativity.

As covered in chapter 3, there are a number of approaches and formalisms that model nonmonotonic reasoning patterns. One interesting point to note is how defeasibility may be encoded on the object-level of a logic. Two examples to illustrate this point are: propositional typicality logic (PTL) [11] represents defeasibility by means of a unary operator $\bullet$ that can be placed before any formula in the language. Another example is defeasible description logics, which has defeasible connectives different from implication, or subsumption as it is referred to in description logics [16]. Compare these two approaches to the KLM framework, where object-level defeasibility is encoded solely in the connective for defeasible implication $\mid\!\sim$. This allows solely for a statement such as $\alpha \mid\!\sim \beta$, which conveys that "α *typically* implies β", or "typical αs are βs" [49]. Whereas the other two approaches allow for expressing defeasible statements on the object-level that cannot be expressed in this framework, for example "all αs are the most typical βs" [11]. How defeasibility is encoded in the language itself is a non-trivial factor in a nonmonotonic logic, and different frameworks can have significantly different expressivity with respect to the defeasible statements that may be represented.

6.2 Summary of Contribution

This dissertation is intended to be a complete overview of the KLM framework for defeasible reasoning, at the time of writing. The primary goal is that this will be useful as a reference for defining defeasible entailment relations, perhaps for an application such as a defeasible reasoning engine, as well as a pedagogical tool that can take any reader from classical propositional logic to rational defeasible entailment using the KLM framework. Since the KLM framework was initially defined, different research groups have built extensions using those foundations. Compiling this work into a single overview should prove useful for future work in this research area, by homogenizing the notation used, while still providing the context for the notation in the initial papers, as well as by providing clarity on subjects that may have proven confusing on first encounter.

6.3 Future Work

Some primary areas of interest for future work in this field include syntax sensitivity, applications in other domains, and explainability.

Syntax sensitivity is the property that the representation of some subset of formulas in a knowledge base affects what is entailed, in the context of syntax-based approaches to entailment [6, 7]. For example, the lexicographic closure is sensitive to the way the knowledge base represents statements. Rational closure, on the other hand, is not at all sensitive, and will entail the same set of inferences regardless of the way the information is expressed, provided the information itself is not changed. Solving syntax sensitivity in presumptive entailments will make these entailment relations more reliable and predictable in behaviour, allowing them to be far more viable in implementations such as reasoning engines.

An area of research that was touched upon in parts of this dissertation is defeasible description logics [16, 21, 27]. While defeasible TBox statements (comparable to propositional knowledge bases) have been defined [15, 18], an ongoing area of research is that of defeasible ABox reasoning [13, 17]. While reasoning with a classical ABox has been defined [13], defining reasoning with respect to a defeasible ABox is an open question [14, 39]. Additionally, there is also the opportunity to do the kind of work of this dissertation in the context of description logics: compiling the various ways KLM-style defeasibility has been incorporated in description logics and providing an overview.

Two active applications of this kind of defeasible reasoning are in the legal domain [41], and in logic programming, specifically using datalog [32, 42]. Legal reasoning is applying formal logic to reasoning about laws and normative systems, and frequently the problem of conflicting obligations arises in various paradoxes [60], and defeasible reasoning has been shown to have success in resolving some such paradoxes [29]. Datalog is a logic programming language that was originally designed as a database querying language, but has found applications in artificial intelligence, notably in RDFox [59] and the DLV system [50]. Datalog has been shown to be extendable to compute the rational closure of a defeasible knowledge base [42], which has implications for how KLM-style defeasible reasoning approaches can be incorporated into various reasoning engines.

Finally, explainable artificial intelligence [43] in this context refer to enriching some reasoning engine with the ability to provide, for all entailments, justifications: a minimal subset of the knowledge base that supports an inference. The goals of explainability are, broadly speaking, to understand entailments that are not obviously inferred by the knowledge, to fix a possibly bugged, or inconsistent, knowledge base, and to gain some better understanding of a knowledge base with which the user may not have prior experience [43]. So far, the majority of work in explainable AI has been in the context of classical reasoning [47], however there is foundational work on extending explainability to the realm of defeasibility [12, 28].

6.4 Conclusion

This dissertation started with classical propositional logic, and then showed how to construct a preferential consequence relation over such a language, with corresponding semantics. Then, the language was extended with a defeasible connective to create a defeasible propositional logic, over which nonmonotonic entailment relations were defined. These entailment relations were then classified into iteratively defined classes. This progression is intended to take the reader from the initial definitions of the KLM framework to the current state. After this dissertation, the reader is encouraged to read the papers defining and extending the KLM framework using this as a reference if necessary.

The aim of this dissertation was to provide a single point of reference that also serves as an overview of the KLM framework. The intention is that using this as a manual of sorts allows for the reader to either manipulate the framework for the purposes of implementing defeasible reasoning in an application, or to assist theoretical extensions and developments of the framework itself.

A concluding message is that every entailment relation necessarily encodes a pattern of reasoning. A knowledge base is essentially a representation of some information, and an entailment relation encodes a reading of the information that determines what type of reasoning is employed. Defeasible entailment relations, therefore, should be viewed not only as classes, defined by postulates describing the associated behaviour, but also as distinct reasoning patterns that may or may not be sensible. No one type of defeasible entailment is a one size fits all, and what type of entailment relation is desirable is domain dependent, comparable to how human reasoning is context dependent. It follows that defeasible entailment is the study of many different relations and their possible applications; not a search for a silver bullet.

Bibliography

[1] Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. The journal of symbolic logic **50**(2), 510–530 (1985)

[2] Alchourrón, C.E., Makinson, D.: On the logic of theory change: Contraction functions and their associated revision functions. Theoria **48**(1), 14–37 (1982)

[3] Avron, A.: Simple consequence relations. Information and Computation **92**(1), 105–139 (1991)

[4] Avron, A.: Gentzen-type systems, resolution and tableaux. Journal of Automated reasoning **10**(2), 265–281 (1993)

[5] Ben-Ari, M.: Mathematical logic for computer science. Springer Science & Business Media (2012)

[6] Benferhat, S., Dubois, D., Prade, H.: How to infer from inconsistent beliefs without revising? In: IJCAI. vol. 95, pp. 1449–1455. Citeseer (1995)

[7] Benferhat, S., Dubois, D., Prade, H.: Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study part 1: The flat case. Studia Logica **58**(1), 17–45 (1997)

[8] Booth, R.: A negotiation-style framework for non-prioritised revision. In: Proceedings of the 8th conference on Theoretical aspects of rationality and knowledge. pp. 137–150. Morgan Kaufmann Publishers Inc. (2001)

[9] Booth, R.: On the logic of iterated non-prioritised revision. In: International Workshop on Conditionals, Information, and Inference. pp. 86–107. Springer (2002)

[10] Booth, R., Hunter, A.: Trust as a precursor to belief revision. Journal of Artificial Intelligence Research **61**, 699–722 (2018)

[11] Booth, R., Meyer, T., Varzinczak, I.: PTL: A propositional typicality logic. In: Logics in Artificial Intelligence, pp. 107–119. Springer (2012)

[12] Brewka, G., Ulbricht, M.: Strong Explanations for Nonmonotonic Reasoning, pp. 135–146. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-22102-7_6, `https://doi.org/10.1007/978-3-030-22102-7_6`

[13] Britz, K., Casini, G., Meyer, T., Moodley, K., Sattler, U., Varzinczak, I.: Rational defeasible reasoning for description logics. Tech. rep., University of Cape Town (2018)

[14] Britz, K., Casini, G., Meyer, T., Moodley, K., Sattler, U., Varzinczak, I.: Theoretical foundations of defeasible description logics. arXiv preprint arXiv:1904.07559 (2019)

[15] Britz, K., Casini, G., Meyer, T., Moodley, K., Varzinczak, I.: Ordered interpretations and entailment for defeasible description logics. Tech. rep., Technical report, CAIR, CSIR Meraka and UKZN, South Africa (2013)

[16] Britz, K., Casini, G., Meyer, T., Varzinczak, I.: Preferential role restrictions (2013)

[17] Britz, K., Casini, G., Meyer, T., Varzinczak, I.: A KLM perspective on defeasible reasoning for description logics. In: Description Logic, Theory Combination, and All That, pp. 147–173. Springer (2019)

[18] Britz, K., Meyer, T., Varzinczak, I.: Semantic foundation for preferential description logics. In: Australasian Joint Conference on Artificial Intelligence. pp. 491–500. Springer (2011)

[19] Britz, K., Meyer, T., Varzinczak, I.: Normal modal preferential consequence. In: Australasian Joint Conference on Artificial Intelligence. pp. 505–516. Springer (2012)

[20] Britz, K., Varzinczak, I.: Defeasible modalities. arXiv preprint arXiv:1310.6409 (2013)

[21] Britz, K., Varzinczak, I.: Introducing role defeasibility in description logics. In: European Conference on Logics in Artificial Intelligence. pp. 174–189. Springer (2016)

[22] Britz, K., Varzinczak, I.: From KLM-style conditionals to defeasible modalities, and back. Journal of Applied Non-Classical Logics **28**(1), 92–121 (2018)

[23] Cadoli, M., Lenzerini, M.: The complexity of propositional closed world reasoning and circumscription. Journal of Computer and System Sciences **48**(2), 255–310 (1994)

[24] Casini, G., Meyer, T.: Belief change in a preferential non-monotonic framework. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. pp. 929–935. International Joint Conferences on Artificial Intelligence (2017)

[25] Casini, G., Meyer, T., Moodley, K., Nortjé, R.: Relevant closure: A new form of defeasible reasoning for description logics. In: Fermé, Eduardoand Leite, J. (ed.) Logics in Artificial Intelligence. pp. 92–106. Springer International Publishing, Cham (2014)

[26] Casini, G., Meyer, T., Varzinczak, I.: Taking defeasible entailment beyond rational closure. In: European Conference on Logics in Artificial Intelligence. pp. 182–197. Springer (2019)

[27] Casini, G., Straccia, U.: Rational closure for defeasible description logics. In: European Workshop on Logics in Artificial Intelligence. pp. 77–90. Springer (2010)

[28] Chama, V., Meyer, T.: Explanation for defeasible entailment (2019)

[29] Chingoma, J., Meyer, T.: Forrester's paradox using typicality (2019)

[30] Darwiche, A., Pearl, J.: On the logic of iterated belief revision. Artificial intelligence **89**(1-2), 1–29 (1997)

[31] Delgrande, J.P., Schaub, T., Jackson], W.K.: Alternative approaches to default logic. Artificial Intelligence **70**(1), 167 – 237 (1994). https://doi.org/https://doi.org/10.1016/0004-3702(94)90106-6, `http://www.sciencedirect.com/science/article/pii/0004370294901066`

[32] Eiter, T., Gottlob, G., Mannila, H.: Disjunctive datalog. ACM Transactions on Database Systems (TODS) **22**(3), 364–418 (1997)

[33] Freund, M.: Preferential reasoning in the perspective of Poole default logic. Artificial Intelligence **98**(1-2), 209–235 (1998)

[34] Gabbay, D.M.: Theoretical foundations for non-monotonic reasoning in expert systems. In: Apt, K.R. (ed.) Logics and Models of Concurrent Systems. pp. 439–457. Springer Berlin Heidelberg, Berlin, Heidelberg (1985)

[35] Gallier, J.H.: Logic for computer science: foundations of automatic theorem proving. Courier Dover Publications (2015)

[36] Geffner, H., Pearl, J.: Conditional entailment: Bridging two approaches to default reasoning. Artificial Intelligence **53**(2-3), 209–244 (1992)

[37] Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.L.: A non-monotonic description logic for reasoning about typicality. Artificial Intelligence **195**, 165–202 (2013)

[38] Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.L.: Semantic characterization of rational closure: From propositional logic to description logics. Artificial Intelligence **226**, 1–33 (2015)

[39] Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.L., et al.: Preferential vs rational description logics: which one for reasoning about typicality?. In: ECAI. pp. 1069–1070 (2010)

[40] Gochet, P., Gribomont, P.: Epistemic logic. In: Handbook of the History of Logic, vol. 7, pp. 99–195. Elsevier (2006)

[41] Grossi, D.: Logic in the law: A concise overview (2011)

[42] Harrision, M., Meyer, T.: Rational preferential reasoning for datalog (2019)

[43] Horridge, M.: Justification based explanation in ontologies. The University of Manchester (United Kingdom) (2011)

[44] Hunter, A., Booth, R.: Trust-sensitive belief revision. In: Twenty-Fourth International Joint Conference on Artificial Intelligence (2015)

[45] Konolige, K.: On the relation between default and autoepistemic logic. Artificial intelligence **35**(3), 343–382 (1988)

[46] Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial intelligence **44**(1-2), 167–207 (1990)

[47] Krötzsch, M., Stepanova, D.: Reasoning Web. Explainable Artificial Intelligence: 15th International Summer School 2019, Bolzano, Italy, September 20–24, 2019, Tutorial Lectures, vol. 11810. Springer Nature (2019)

[48] Lehmann, D.: Another perspective on default reasoning. Annals of Mathematics and Artificial Intelligence **15**(1), 61–82 (1995)

[49] Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial intelligence **55**(1), 1–60 (1992)

[50] Leone, N., Pfeifer, G., Faber, W., Eiter, T., Gottlob, G., Perri, S., Scarcello, F.: The dlv system for knowledge representation and reasoning. ACM Trans. Comput. Logic **7**(3), 499–562 (Jul 2006). https://doi.org/10.1145/1149114.1149117, `https://doi.org/10.1145/1149114.1149117`

[51] Magnus, P.D.: Forall X: An Introduction to Formal Logic (2005)

[52] McCarthy, J.: Circumscription—a form of non-monotonic reasoning. Artificial Intelligence **13**(1-2), 27–39 (1980)

[53] McCarthy, J.: Applications of circumscription to formalizing common sense knowledge. Artificial Intelligence **28**, 89–116 (1986)

[54] McDermott, D., Doyle, J.: Nonmonotonic logic I. Artificial Intelligence **13**, 41–72 (1980)

[55] McDermott, D.: Nonmonotonic logic II: Nonmonotonic modal theories. Journal of the ACM (JACM) **29**(1), 33–57 (1982)

[56] Monk, J.D.: The contributions of Alfred Tarski to algebraic logic. The Journal of Symbolic Logic **51**(4), 899–906 (1986), `http://www.jstor.org/stable/2273903`

[57] Moore, R.C.: Possible-world semantics for autoepistemic logic. Tech. rep. (1984)

[58] Moore, R.C.: Semantical considerations on nonmonotonic logic. Artificial Intelligence **25**(1), 75–94 (1985)

[59] Nenov, Y., Piro, R., Motik, B., Horrocks, I., Wu, Z., Banerjee, J.: RDFox: A Highly-Scalable RDF Store. In: Arenas, M., Corcho, O., Simperl, E., Strohmaier, M., d'Aquin, M., Srinivas, K., Groth, P., Dumontier, M., Heflin, J., Thirunarayan, K., Staab, S. (eds.) The Semantic Web - ISWC 2015. pp. 3–20. Springer International Publishing, Cham (2015)

[60] Parent, X., van der Torre, L.: Detachment in normative systems: Examples, inference patterns, properties. IfCoLog Journal of Logics and their Applications **4**(9), 2295–3039 (2017)

[61] Reiter, R.: A logic for default reasoning. Artificial Intelligence **13**(1-2), 81–132 (1980)

[62] Reiter, R., Criscuolo, G.: Some representational issues in default reasoning. Computers & Mathematics with Applications **9**(1), 15–27 (1983)

[63] Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach. Pearson Education Limited, (2016)

[64] Shoham, Y.: A Semantical Approach to Nonmonotonic Logics, p. 227–250. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1987)

[65] Shoham, Y.: Nonmonotonic logics: Meaning and utility. In: IJCAI. vol. 10, pp. 388–393. Citeseer (1987)

[66] Shoham, Y.: Reasoning about change: Time and causation from the standpoint of artificial intelligence. (1988)

[67] Tarski, A.: On the concept of logical consequence. Logic, semantics, metamathematics **2**, 1–11 (1936)

[68] Von Wright, G.H.: Deontic logic. Mind **60**(237), 1–15 (1951)

[69] Wasilewska, A.: Logics for Computer Science: Classical and Non-Classical (01 2018). https://doi.org/10.1007/978-3-319-92591-2