Enriching deontic logic with typicality

By

Julian Chingoma

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

Declaration

I, Julian Chingoma, do hereby declare that this Masters thesis titled "Enriching deontic logic with typicality" is my own work and has not being 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: 19 June 2020

Acknowledgements

Firstly, I wish to thank my supervisor, Thomas Meyer, whose guidance and encouragement has been pivotal in not only this particular research, but my entire academic journey thus far. This work would not have been possible without his support.

I would also like to thank the Centre for Artificial Intelligence (CAIR) for their financial support. The funding they provided made the undertaking of this research significantly less challenging.

I am grateful for the contributions of Leon van der Torre, Giovanni Casini and Ivan Varzinczak, whose teachings during different stages of the dissertation's development have been valuable to its completion.

I also extend a thanks to my fellow students as well as the staff at the University of Cape Town. They have made conducting this research an enjoyable process throughout.

Finally, I would like to thank my family and friends who have supported me during every stage of the process. Their patience and encouragement have been crucial throughout my studies.

Enriching deontic logic with typicality

Julian Chingoma

Abstract

Legal reasoning is a method that is applied by legal practitioners to make legal decisions. For a scenario, legal reasoning requires not only the facts of the scenario but also the legal rules to be enforced within it. Formal logic has long been used for reasoning tasks in many domains. Deontic logic is a logic which is often used to formalise legal scenarios with its built-in notions of obligation, permission and prohibition. Within the legal domain, it is important to recognise that there are many exceptions and conflicting obligations. This motivates the enrichment of deontic logic with not only the notion of defeasibility, which allows for reasoning about exceptions, but a stronger notion of typicality which is based on defeasibility. KLM-style defeasible reasoning introduced by Kraus, Lehmann and Magidor (KLM), is a logic system that employs defeasibility while a logic that serves the same role for the stronger notion of typicality is Propositional Typicality Logic (PTL). Deontic paradoxes are often used to examine deontic logic systems as the scenarios arising from the paradoxes' structures produce undesirable results when desirable deontic properties are applied to the scenarios. This is despite the various scenarios themselves seeming intuitive. This dissertation shows that KLM-style defeasible reasoning and PTL are both effective when applied to the analysis of the deontic paradoxes. We first present the background information which comprises propositional logic, which forms the foundation for the other logic systems, as well as the background of KLM-style defeasible reasoning, deontic logic and PTL. We outline the paradoxes along with their issues within the presentation of deontic logic. We then show that for each of the two logic systems we can intuitively translate the paradoxes, satisfy many of the desirable deontic properties and produce reasonable solutions to the issues resulting from the paradoxes.

Contents

Declaration **i**

Acknowledgements **ii**

Abstract **iii**

1 Introduction **1**
- 1.1 Research Objectives . . . 3
- 1.2 Dissertation outline . . . 4

2 Propositional Logic **6**
- 2.1 Language . . . 6
- 2.2 Semantics . . . 8

3 KLM-style Defeasible Reasoning **10**
- 3.1 KLM approach . . . 10
- 3.2 Language . . . 11
- 3.3 Semantics . . . 12
- 3.4 Properties of rational defeasible entailment relation . . . 12
- 3.5 Rational Closure . . . 15
- 3.6 Lexicographic Closure . . . 18

4 Deontic Logic **22**
- 4.1 Language . . . 23
- 4.2 Semantics . . . 23
- 4.3 Deontic Properties . . . 24
- 4.4 Paradoxes . . . 28
  - 4.4.1 Forrester's paradox . . . 28
  - 4.4.2 Chisholm's paradox . . . 29

4.4.3 Fence paradox . . . 31
4.4.4 Trump/Kim paradox . . . 32
4.4.5 Van Fraassen's paradox . . . 33

5 Propositional Typicality Logic 35
5.1 Language . . . 35
5.2 Semantics . . . 36
5.3 LM-entailment . . . 36
5.4 PT-entailment . . . 39

6 Deontic Analysis Process 42

7 Deontic Analysis Using KLM-style Defeasible Reasoning 44
7.1 Representation . . . 44
7.2 Properties . . . 45
7.2.1 Ought Implies Can Principle and Violations . . . 45
7.2.2 Remaining properties . . . 47
7.2.3 Strengthening of the Antecedent . . . 48
7.3 Why use lexicographic closure? . . . 49
7.4 Paradoxes . . . 50
7.4.1 Forrester's paradox . . . 50
7.4.2 Chisholm's paradox . . . 53
7.4.3 Fence paradox and exceptions . . . 55
7.4.4 Trump/Kim paradox . . . 56
7.4.5 Van Fraassen's paradox . . . 57
7.5 Conclusion . . . 58

8 Deontic Analysis Using LM-entailment 59
8.1 Representation . . . 59
8.1.1 Bullet operator on both sides? . . . 60
8.1.2 Exceptions . . . 61
8.1.3 Representation summary . . . 62
8.2 Properties . . . 63
8.2.1 Ought Implies Can Principle . . . 63
8.2.2 Restricted Strengthening of the Antecedent . . . 64
8.2.3 Weakening . . . 67
8.2.4 Deontic Detachment . . . 68
8.2.5 Factual Detachment . . . 69

8.2.6 Distribution 70
8.2.7 Conjunction 70
8.2.8 Disjunction 71
8.2.9 Transitivity 72
8.2.10 Conclusions 73
8.3 Paradoxes 74
8.3.1 Forrester's paradox 74
8.3.2 Chisholm's paradox 75
8.3.3 Fence paradox 77
8.3.4 Trump/Kim paradox 78
8.3.5 Van Fraassen's paradox 79
8.4 Conclusion 80

9 Deontic Analysis Using PT-entailment **82**
9.1 Representation 82
9.2 Properties 83
9.3 Paradoxes 83
9.3.1 Contrary-to-duty example 83
9.3.2 Extended contrary-to-duty example 84
9.3.3 Exception example 85
9.4 Conclusion 86

10 Conclusions And Future Work **87**
10.1 Conclusions 87
10.2 Future Work 89

Chapter 1

Introduction

Legal reasoning is a method used when one is given normative knowledge and facts and must produce legal decisions based on the information presented and the application of certain rules applied to this information [16]. We do not limit our interest to legal reasoning, but we want to move beyond the legal domain into the general reasoning about regulations. Logic has for a long time been used to formalise legal norms and study legal reasoning [9]. Deontic logic is one such field of logic which has been used to work with normative reasoning tasks [21, 23]. The difference between "what is the case" and "what should be the case" is fundamental to law and this naturally translates to deontic logic with its handling of normative concepts such as obligation, permission and prohibition. This represents the principal motivation for investigating deontic logic, as it can intuitively formalise normative scenarios such as:

1. It is obligatory that a person pay tax
2. John is a person
3. John did not pay tax

This is a plausible scenario in which John violated the obligation that one must pay their taxes and it represents a breaking of the law in the actual world. Since deontic logic can naturally represent such cases, it is worth studying this logic when investigating legal reasoning. Research in deontic logic often focuses on how to express scenarios using the language of a chosen deontic logic variation [12]. An interest in building computing tools for regulatory decision making motivates to not merely look at the expressive capabilities of deontic logic but also the reasoning capabilities. And generally when performing reasoning in the actual world, we come across exceptional cases that conflict with our usual way of thinking. This occurs in the legal domain often, so we want our deontic logic to handle such occurrences. Therefore, the focus of this research will be to find a way of introducing a concept called defeasibility into a deontic setting. Defeasibility allows for reasoning about exceptions in a domain, distinguishing between "what is usually the case" and "what is actually the case" [3, 4, 27]. The following is a situation which contains an exception and we cannot reasonably deal with it in a logic without defeasibility:

1. Students do not pay tax
2. Students who are part-time workers pay tax
3. John is a student who pays tax

This scenario provides an issue since the first two statements contradict each other therefore John cannot exist, as a student who pays tax cannot exist even if they are a part-time worker. This is not intuitive, as we should be able to conclude from the three statements that John is a student who is also a part-time worker. And that conclusion is in line with the usual reasoning that humans employ, as some exceptions and conflicts are presented in everyday life that require the retraction of previous conclusions. The statements of the above example with an exception can instead, be stated as follows if using a logic equipped with defeasibility:

1. Students usually do not pay tax
2. Students who are part-time workers usually pay tax
3. John is a student who pays tax

This is a representation that does not produce inconsistencies and can be reasoned upon with the various reasoning algorithms available for defeasible reasoning. The first two statements now provide the room for exceptional situations, as is the case with John in our example. Intuitively, the above tells us that John is a part-time worker and there is only a problem if we learn that John, in fact, is not a part-time worker. Now we motivate the enriching of deontic logic with this notion of defeasibility. Since deontic logic is the tool of choice when looking to formalise the reasoning of laws and regulations, it is important to note that there are already notions of defeasibility in the world of legal reasoning. The introduction of new information/regulations can cause laws to conflict and/or present exceptions which make existing laws inapplicable [9]. The following is an example where such issues occur. Assume the following statements to be true:

1. Students must not pay tax
2. Students who are part-time workers must pay tax

One can represent these obligations using deontic logic but when presented with the fact that there is a student who works part-time and must, therefore, pay taxes, it presents a violation of the first obligation when there intuitively should not be. The law intuitively means that a typical student should not pay tax while part-time working students are not typical students but are an exception to the norm rather than a violation of the first obligation. So it seems reasonable to desire a deontic logic with the ability to cater for exceptions in order to reason about such normative scenarios.

1.1 Research Objectives

This section aims to give a broad outline of the research objectives without delving into significant detail. First, we look to determine how effective defeasibility is when applied in a deontic environment. This will start by investigating the extent to which the deontic statements can be represented using the following logic system. The logic often referred to as KLM-style defeasible reasoning is the logic which represents defeasibility, which we will look at during the dissertation [4, 7, 13]. We will begin by looking at how we can represent deontic statements using KLM-style defeasible reasoning. With this approach comes two well-established reasoning methods, so we aim to investigate the extent to which they are useful in a deontic environment. Beyond defeasibility, we are also interested in the stronger notion of typicality. Typicality is based on defeasibility and is a notion used in Propositional Typicality Logic (PTL), where its extra expressive power makes it a more powerful version of the defeasibility presented by KLM-style defeasible reasoning [1].

The first part, dealing with the KLM approach, will inform us on whether it is worth examining the introduction of typicality into a deontic environment. This examination of PTL's suitability in a deontic setting will require looking at some of the following issues. Is there an intuitive way to represent deontic statements using the language of PTL? Similarly to defeasible reasoning, PTL has multiple reasoning methods which can be investigated. The two different PTL reasoning methods seem to produce reasonable conclusions for given PTL statements, but neither has been tested in specific domains such as the one of legal reasoning. We ask if either of the PTL entailment methods produce reasonable conclusions when applied to the realm of deontic reasoning.

1.2 Dissertation outline

This dissertation can be logically split into two parts. The first largely deals with presenting the background information required to conduct the analysis that is found in the second part. Chapter 2 introduces propositional logic as this forms the foundation of the various logic systems we analyse in the research. This chapter outlines the propositional language and thereafter, the notion of valuations is introduced as this is a key part of the semantics of our logic systems of interest. Chapter 3 presents the KLM-style defeasible reasoning approach. This is a logic which extends propositional logic with the introduction of an operator to represent defeasibility. Its language is briefly described and subsequently, properties are given which tell us what we can conclude from the sets of KLM-style statements. The chapter will end with the presentation of the KLM semantics and two reasoning methods that can be used when dealing with KLM-style statements. Chapter 4 presents a specific deontic logic system which is also an enrichment of propositional logic, with the inclusion of an operator to represent obligations. Its language is briefly described, and its semantics are presented thereafter. We then detail some deontic properties which are generally deemed to be desirable and then discuss some of the famous deontic paradoxes which have interested researchers in the field. Propositional Typicality Logic (PTL) is then detailed in chapter 5 in a similar fashion to the previous logic systems. Once again, there is a brief description of the language as PTL also has propositional logic as its foundation along with the addition of a typicality operator. The semantics for PTL are presented and two of the PTL reasoning methods are given. Chapter 6 will outline the research processes we employed and after this chapter is where the dissertation transitions into the analysis part. The use of KLM-style defeasibility on deontic scenarios is explored in chapter 7. We first show how we chose to represent deontic scenarios using KLM-style statements and proceed to investigate what we can conclude when using this logic system on the deontic paradoxes. Chapters 8 and 9 each lay out the analysis of one of PTL's two reasoning methods in a similar vein to chapter 7. In chapter 8, how deontic statements will be formally represented using PTL will be detailed first. We then examine what we can derive using the one reasoning method. Chapter 9, on the other hand, will jump straight to the analysis of the conclusions, as the representation from chapter 8 is carried over into this chapter. The reasons for this are communicated at the beginning of chapter 9. In the final chapter of the dissertation, chapter 10, we present our conclusions. This chapter is where we briefly restate the work of the dissertation, consolidate the results from the prior three chapters and come to a final determination on the effectiveness of the various approaches. Chapter 10 ends with a discussion on potential avenues for future work.

Chapter 2

Propositional Logic

This chapter serves to present propositional logic, the logical foundation of the three logic systems which we examine during this dissertation. Propositional logic is utilised to formalise statements that can considered to be either true or false, and is used to study the logical relationships between these statements [12]. The formal language, which is presented below, consists of propositional letters along with operators and constants which can be used to form propositional formulas. These formulas are what formally represent the statements that are to be studied. The propositional logic semantics, which give the formal meaning for the given statements, are also shown and the important notions of satisfaction and entailment in propositional logic are also defined.

2.1 Language

We now specify the language of propositional logic which determines the type of statements that can be formulated using propositional logic.

Definition 2.1. *The propositional language* $\mathcal{L}$ *is formed by a set of propositional letters* Φ*, such as* p*,*q *and* r*, together with the following operators and constants* [6, 12, 25]*:* $\{\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \bot, \top\}$.

These operators and constants can be used with the propositional letters to recursively build propositional formulas. These formulas are usually denoted by α, β, γ,etc. The following are examples of some propositional formulas that can be created, where $\alpha = p$ and $\beta = q$: $\{\alpha, \beta, \neg\alpha, \alpha\wedge\beta, \alpha\vee\beta, \alpha \rightarrow \beta, \alpha \leftrightarrow \beta, \alpha \rightarrow \bot, \top \rightarrow \beta\}$.

Propositional letters such as p and q can represent any proposition and for this example, they can be read as *"John is a student"* and *"John pays tax"* respectively. The operators and constants of $\mathcal{L}$ are now given in further detail.

- $\neg$ | this is an operator which represents negation, so $\neg p$ is read as *"not p"* or *"John is not a student"*.
- $\wedge$ | this is an operator which represents conjunction, so $p \wedge q$ is read as *"p and q"* or *"John is a student and John pays tax"*.
- $\vee$ | this is an operator which represents disjunction, so $p \vee q$ is read as *"p or q"* or *"John is a student or John pays tax"*.
- $\rightarrow$| this is an operator used to represent implication and what logically follows from a premise. $p \rightarrow \neg q$ would translate to *"p implies* $\neg q$*"* or *"If John is a student then John does not pay tax"*.
- $\leftrightarrow$| this is an operator used to represent the double implication/"if and only if". $p \leftrightarrow \neg q$ would translate to *"p implies* $\neg q$ *and* $\neg q$ *implies p"* or *"John is a student if and only if John does not pay tax"*.
- $\perp$ | this is a constant which represents a contradiction. This can be thought of as *"p is true and p is false"*/$p \wedge \neg p$, which cannot be true. This could be represented as the statement *"John is a student and John is not a student"*.
- $\top$ | this constant which represents tautology or absolute truth. This is the opposite of $\perp$ and can be represented by *"p is true or p is false"*/$p \vee \neg p$, which is always true. This could be represented as the statement *"John is a student or John is not a student"*.

Since the reasoning aspect is of interest to us, it is important to detail the notion of entailment, with entailment being the concept of what conclusions logically follow from a set of premises [1]. The classical way to do this in propositional logic is to look at the truth assignments of the propositional letters, usually denoted using valuations. A valuation is an assignment of either true or false to each propositional letter. For example, the valuation $\{p, q\}$ states that both p and q are true while $\{p, \neg q\}$ is a valuation where p is true and q is false.

2.2 Semantics

With the language defined, we now detail the propositional logic semantics. The semantics tell us what we can conclude about the truth of a propositional formula.

Definition 2.2. *Let* W *denote the set of all possible valuations. The notion of satisfaction of a propositional formula is commonly denoted using* $\models$*. Given a valuation* $s \in W$*, a propositional letter* p *and propositional formulas* α *and* β*, we can define satisfaction in the propositional language as follows [6, 22, 25]:*

- $s \models p$ *iff* p *is true in the valuation* s
- $s \models \neg\alpha$ *iff not* $s \models \alpha$*, as in* α *is false in* s
- $s \models \alpha \wedge \beta$ *iff* $s \models \alpha$ *and* $s \models \beta$*, as in* α *and* β *are both true in* s
- $s \models \alpha \vee \beta$ *iff* $s \models \alpha$ *or* $s \models \beta$*, as in at least one of* α *and* β *hold in* s
- $s \models \alpha \rightarrow \beta$ *iff* $s \models \neg\alpha \vee \beta$*, as in at least one of* $\neg\alpha$ *and* β *hold in* s
- $s \models \alpha \leftrightarrow \beta$ *iff* $s \models \alpha \rightarrow \beta$ *and* $s \models \beta \rightarrow \alpha$

This satisfaction definition can be applied to any propositional formula formed by the propositional letters, operators and constants. A knowledge base is a set of statements that represent information about some domain. A classical knowledge base, let's say $\mathcal{KB}$, is a set of propositional formulas of the propositional logic language. Classical entailment tells us what logically follows from a knowledge base and classical entailment relies upon the above concept of satisfaction of propositional formulas. If all the valuations that model a knowledge base $\mathcal{KB}$, which refers to satisfying all the formulas in $\mathcal{KB}$, also satisfy a formula, let's say α, then we say that $\mathcal{KB}$ entails α. This is given in the following definitions.

Definition 2.3. *We say that a valuation, let's say* $s \in W$*, satisfies a knowledge base* $\mathcal{KB}$*,* $s \models \mathcal{KB}$*, if and only if* $s \models \alpha$ *for every propositional formula* $\alpha \in \mathcal{KB}$*.*

Definition 2.4. *Let* $\mathcal{KB}$ *be a knowledge base and* α *be a propositional formula. We say that* $\mathcal{KB}$ *entails* α*,* $\mathcal{KB} \models \alpha$*, if and only if for every* $s \in W$ *such that* $s \models \mathcal{KB}$*, we have that* $s \models \alpha$*.*

As seen in definition 2.4, the $\models$ symbol is not only used to denote satisfaction but is also employed to represent classical entailment, which is entailment from a classical knowledge base. Again, we denote the entailment of a propositional formula α from $\mathcal{KB}$ as $\mathcal{KB} \models \alpha$, so for an example formula $p \rightarrow q$, $\mathcal{KB} \models p \rightarrow q$ asserts that we can derive $p \rightarrow q$ from a knowledge base $\mathcal{KB}$. For the statement $p \rightarrow q$ to follow from a knowledge base, then we check for every valuation that satisfies the knowledge base, if we have that p is true in such a valuation then q must also be true in that valuation. Another way to think about these valuations is when we have statements that we know to be true and must determine which valuations satisfy these statements. So for example, let's take the propositional letters p and q, and we have the valuation $\{p, \neg q\}$. We are able to conclude that the valuation entails $p \vee q$ but not $p \wedge q$.

Chapter 3

KLM-style Defeasible Reasoning

This chapter presents the KLM approach to defeasible reasoning. We begin by introducing the language and semantics of this logic system. We then outline some properties which inform us about the entailment of some of the KLM approach's reasoning methods. Two of these reasoning methods are then outlined with their differences briefly discussed. We also present their two respective algorithms which can be used to perform the reasoning tasks.

3.1 KLM approach

The first logic we investigate is an enrichment of propositional logic and is a form of non-monotonic reasoning, which is a form of reasoning that allows for conclusions to be retracted. The property of monotonicity is one that classical logic systems satisfy which states that the addition of new information strictly leads to more conclusions. This does not align with the usual human thinking which better handles exceptions. The use of this non-monotonic reasoning system allows us to reason defeasibly. This logic system is the KLM approach to defeasible reasoning. It is a logic that was proposed by Kraus, Lehmann and Magidor, hence the name KLM [13]. Before looking at the KLM-style defeasible reasoning in detail, it is important to acknowledge that this is not the only option to perform defeasible reasoning. Other approaches include circumscription, default logic and auto-epistemic logic [18, 19, 24]. Since there are other non-monotonic reasoning approaches, it seems reasonable to justify the use of the KLM approach.

The first benefit to using the KLM approach is that it is systematic. This tells us that it has set, formal requirements that are to be satisfied for the system to produce intuitively reasonable results when dealing with non-monotonic reasoning scenarios [15]. We will see this with the properties that are detailed later in this chapter. Another factor considered is that the computational complexity of the system measures well in comparison to other non-monotonic approaches [3], this refers particularly to the reasoning methods. There is also the added benefit that the process of reasoning with the system is automatic and therefore does not require considerable input from users. This contrasts some other approaches which require much more user-involvement at various stages. Lastly, this approach can be reduced to classical reasoning which means that we can use well-established methods and software tools for reasoning scenarios [2, 3]. All these mentioned benefits would be especially helpful for an individual building software tools for defeasible reasoning. We now proceed to define the formal language and ranked-interpretation semantics of the KLM approach.

3.2 Language

The following is the definition of the language of KLM-style defeasible reasoning. The KLM approach's language is an enrichment of the propositional logic language.

Definition 3.1. *The language of KLM-style defeasible reasoning is formed by a set of propositional letters Φ, the operators and constants of propositional logic with the addition of the following defeasible implication operator, $\mid\!\sim$. Thus the KLM approach's language is an enrichment of the propositional logic language with defeasible implications of the form $\alpha \mid\!\sim \beta$ where α and β are propositional formulas.*

The defeasible implication operator $\mid\!\sim$ is the defeasible version of the classical implication $\rightarrow$. Defeasible implications represent implications that we can potentially reject in exceptional circumstances and are read as "α typically implies β" as opposed to statements made with the classical implication.

3.3 Semantics

With the language for the KLM approach defined, we proceed to detail the semantics. The semantics for the KLM approach are defined in terms of ranked interpretations.

Definition 3.2. *A ranked interpretation, $\mathscr{R}$, is a set of valuations, let's say $V \subseteq W$ where W denotes the set of all possible valuations, along with a binary relation $\leq$ where the valuations are ranked with some valuations being preferred to others. $\leq$ is a modular order and holds the properties of reflexivity, transitivity and anti-symmetry. Since we have anti-symmetry then valuations that are of the same rank or level of preference are incomparable. Incomparability states that if the valuations, let's say v and w, are of the same rank/level, then we have neither $v \leq w$ or $w \leq v$. Given a ranked interpretation $\mathscr{R}$ and a formula α, the set of valuations that satisfy α are represented as $[\![\alpha]\!]^{\mathscr{R}}$, where $[\![\alpha]\!]^{\mathscr{R}} = \{v \in V \mid v \models \alpha\}$ [1, 13]. We say that a defeasible implication, let's say $\alpha \mid\!\sim \beta$, is satisfied in a ranked interpretation if the minimal valuations where α is true are valuations where β is true, $min_{\leq}[\![\alpha]\!]^{\mathscr{R}} \subseteq [\![\beta]\!]^{\mathscr{R}}$ [1, 13, 15]. That is to say that given $[\![\alpha]\!]^{\mathscr{R}}$, $\forall v' \in V$, if $v' \in \{v \in [\![\alpha]\!]^{\mathscr{R}} : \nexists\, w \in [\![\alpha]\!]^{\mathscr{R}},$ such that $w < v\}$, then $v' \models \beta$.*

3.4 Properties of rational defeasible entailment relation

We now present the defeasible counterpart to the classical entailment relation that we have seen previously, $\models$. We will use this defeasible entailment relation to denote what we can derive from a defeasible knowledge base.

Definition 3.3. *A defeasible knowledge base is a set which includes classical propositional logic formulas, such as $\neg\alpha$, $\alpha \vee \beta$ and $\alpha \rightarrow \beta$, along with defeasible implications such as $\alpha \mid\!\sim \beta$.*

Defeasible entailment will be denoted using the relation $\mid\!\approx$ and for example, a statement such as $\mathcal{KB} \mid\!\approx \alpha \mid\!\sim \beta$ tells us that $\alpha \mid\!\sim \beta$ defeasibly follows from the knowledge base. Below is a list of properties that the defeasible entailment relation, $\mid\!\approx$, ought to satisfy in order to be considered rational. These properties are to be interpreted as if we have a defeasible knowledge base, $\mathcal{KB}$, which makes use of propositional formulas such as α, β and γ.

The following is the derivation of one of these properties, the Left Logical Equivalence property. This derivation also serves as an example to illustrate the structure of the remaining property derivations. The general form of the derivations will be as follows. The premises will appear first and these will tell us what the knowledge base currently entails, either classically or defeasibly. Then if there is a separating line, then what is below that line will be what logically follows from the premises.

$$\frac{\models \alpha \leftrightarrow \beta, \mathcal{KB} \approx\!\!\!\!| \;\, \alpha \mid\!\sim \gamma}{\mathcal{KB} \approx\!\!\!\!| \;\, \beta \mid\!\sim \gamma}$$

The above derivation tells us that given a defeasible knowledge base $\mathcal{KB}$, we have that $\alpha \leftrightarrow \beta$ is a tautology and $\alpha \mid\!\sim \gamma$ follows from $\mathcal{KB}$. The Left Logical Equivalence property states that $\beta \mid\!\sim \gamma$ logically follows from the knowledge base. We now give a representative example to demonstrate the intuition of the property. If we have that *“being a student is equivalent to being a young person”* and *“being a student usually implies having a student card”* then is it reasonable to derive that *“being a young person usually implies having a student card”*. Similarly, illustrative examples will be given for the remaining properties as they are detailed. The following is a list of the remaining properties along with a brief example to illustrate their intuition.

Reflexivity $\mathcal{KB} \approx\!\!\!\!| \;\, \alpha \mid\!\sim \alpha$

This tells us that $\alpha \mid\!\sim \alpha$ follows from the knowledge base and this applies for every propositional formula in $\mathcal{KB}$. We could read this as *"being a student usually implies being a student"*

Cumulative Transitivity $\dfrac{\mathcal{KB} \approx\!\!\!\!| \;\, \alpha \mid\!\sim \beta, \mathcal{KB} \approx\!\!\!\!| \;\, \alpha \wedge \beta \mid\!\sim \gamma}{\mathcal{KB} \approx\!\!\!\!| \;\, \alpha \mid\!\sim \gamma}$

If we have that both $\alpha \mid\!\sim \beta$ and $\alpha \wedge \beta \mid\!\sim \gamma$ follow from the knowledge base, this property tells us that we can derive $\alpha \mid\!\sim \gamma$ from the knowledge base. If we have that *“being a lecturer usually implies being a staff member”* and we also have that *“being a lecturer and being a staff member usually implies having a staff number”* then we should be able to derive *“being a lecturer usually implies having a staff number”*.

Conjunction

$$\frac{\mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta, \mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \gamma}{\mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta \wedge \gamma}$$

If have that both $\alpha \mathrel{|\!\sim} \beta$ and $\alpha \mathrel{|\!\sim} \gamma$ follow from the knowledge base, this property tells us that we can conclude that $\alpha \mathrel{|\!\sim} \beta \wedge \gamma$ follows from the knowledge base. If we have for example that *“being a student usually implies having a student card”* and also have *“being a student usually implies paying the fees in full”* then we would want it to be the case that *“being a student usually implies having a student card and paying the fees in full”*.

Disjunction

$$\frac{\mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \gamma, \mathcal{KB} \mathrel{|\!\approx} \beta \mathrel{|\!\sim} \gamma}{\mathcal{KB} \mathrel{|\!\approx} \alpha \vee \beta \mathrel{|\!\sim} \gamma}$$

If we have that both $\alpha \mathrel{|\!\sim} \gamma$ and $\beta \mathrel{|\!\sim} \gamma$ follow from the knowledge base, this property tells us that we can conclude that $\alpha \vee \beta \mathrel{|\!\sim} \gamma$ follows from the knowledge base. If we have that *“being a student entering the library usually implies having a student card”* and we also have *“being a student entering the computer lab usually implies having a student card”* then we should be able to derive *“being a student entering the library or being a student entering the computer lab usually implies having a student card”*.

Weakening

$$\frac{\models \beta \rightarrow \gamma, \mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta}{\mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \gamma}$$

If we have $\beta \rightarrow \gamma$ as a tautology and that $\alpha \mathrel{|\!\sim} \beta$ follows from the knowledge base, this property tells us that we can conclude that $\alpha \mathrel{|\!\sim} \gamma$ follows from the knowledge base. If we have for example that *“being a student implies having a student card”* and *“paying the fess in full usually implies being a student”* then we should be able to derive *“paying the fess in full usually implies having a student card”*.

Cautious Monotonicity

$$\frac{\mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \beta, \mathcal{KB} \mathrel{|\!\approx} \alpha \mathrel{|\!\sim} \gamma}{\mathcal{KB} \mathrel{|\!\approx} \alpha \wedge \beta \mathrel{|\!\sim} \gamma}$$

If we have that both $\alpha \mathrel{|\!\sim} \beta$ and $\alpha \mathrel{|\!\sim} \gamma$ follow from the knowledge base, this property tells us that we can conclude that $\alpha \wedge \beta \mathrel{|\!\sim} \gamma$ follows from the knowledge base. If we have for example that *“being a student usually implies* *having a student card"* and also have *"being a student usually implies paying the fees in full"* then we would want it to be the case that *"being a student and having a student card usually implies paying the fees in full"*.

Rational Monotonicity $$\frac{\mathcal{KB} \approx \alpha \mid\!\sim \beta, \mathcal{KB} \not\approx \alpha \mid\!\sim \neg\gamma}{\mathcal{KB} \approx \alpha \wedge \gamma \mid\!\sim \beta}$$

If we have that $\alpha \mid\!\sim \beta$ follows from the knowledge base and also have that $\alpha \mid\!\sim \neg\gamma$ does not follow from the knowledge base, this property tells us that we can conclude that $\alpha \wedge \gamma \mid\!\sim \beta$ follows from the knowledge base. If we have for example that *"being a student usually implies having a student card"* and also have *"being a student does not usually imply not paying the fees in full"* then we would want it to be the case that *"being a student and paying the fees in full usually implies having a student card"*.

An entailment relation that satisfies all the above properties is called a rational entailment relation. The semantics of these rational entailment relations of the KLM approach is defined by ranked interpretations [15].

3.5 Rational Closure

Rational closure is the first of two KLM reasoning methods that we examine. The rational closure algorithm will construct a ranking where every formula has a rank. Determining the rank of a formula requires one to look at how exceptional it is and will require the definition of $\overrightarrow{\mathcal{KB}}$ which is the materialisation set of the knowledge base $\mathcal{KB}$. A materialisation being the converting of defeasible implications into classical implications so we have $\overrightarrow{\mathcal{KB}} = \{\alpha \rightarrow \beta \mid \alpha \mid\!\sim \beta \in \mathcal{KB}\}$. A formula α is exceptional in a defeasible knowledge base, let's say, $\mathcal{KB}$ if and only if $\overrightarrow{\mathcal{KB}} \models \neg\alpha$ [7, 15]. This tells us that the negation of the formula is classically entailed by the materialisation set. And a defeasible implication is exceptional if its antecedent is exceptional with respect to $\mathcal{KB}$. For example, $\alpha \mid\!\sim \beta$ is exceptional in $\mathcal{KB}$ if α is exceptional in $\mathcal{KB}$. To determine the degree to which a defeasible implication is exceptional, we construct a non-increasing sequence of exceptional subsets of $\mathcal{KB}$. We say that $\mathcal{E}(\mathcal{KB})$ is a set of the exceptional defeasible implications of $\mathcal{KB}$ [7, 15]. Now, we consider a sequence of subsets of $\mathcal{KB}$, $\mathcal{C}_i$ for $i > 0$, where $\mathcal{C}_0 = \mathcal{KB}$, and $\mathcal{C}_i = \mathcal{E}(\mathcal{C}_{i-1})$. For a $\mathcal{KB}$, there is an $n \geq 0$ such that $\mathcal{C}_n = \emptyset$ or for all $m > n$, $\mathcal{C}_m = \mathcal{C}_n$. We then say that the rank of a formula, let's say α, will be the smallest natural number, i, in the subset sequence such that α is not exceptional. If the formula is exceptional for all the subsets in the sequence then it has infinite rank and also note that classical statements are also given an infinite rank. The steps for the rational closure algorithm are now listed.

Step 1 Firstly, obtain a ranking for the knowledge base's materialisation set, $\overrightarrow{\mathcal{KB}}$, based on exceptionality. $(R_0, ..., R_{n-1}, R_\infty, n)$ will be the ordered tuple which represents the rankings.
Step 2 If we now ask if the defeasible implication $\alpha \mid\!\sim \beta$ can be derived by the knowledge base. We must first check if $\neg\alpha$ can be derived from the rankings $(R_0, ..., R_{n-1}, R_\infty, n)$.
Step 3 If $\neg\alpha$ is derivable than the group of the least exceptional defeasible implications will be removed from the ranking and no longer be considered. We then return to step 2. If $\neg\alpha$ cannot be derived then proceed to step 4.
Step 4 Once $\neg\alpha$ is no longer derivable from the rankings, then we can check if $\alpha \rightarrow \beta$ is derivable from the rankings.

To present the rational closure algorithm, we will use an example to better illustrate the steps taken. we take a set $\{b \mid\!\sim f, b \mid\!\sim w, p \mid\!\sim \neg f, p \rightarrow b, r \rightarrow b\}$ where p is a *"penguin"*, b is a *"bird"*, r is a *"robin"*, f is *"to fly"* and w is *"has wings"*. The knowledge base classically entails $\neg p$ because we have that *"birds usually fly"*, *"penguins do not usually fly"* and *"penguins are birds"*. The logical consequence of this is that there are no penguins since the existence of penguins will cause a conflict. Since the materialisation of this set classically entails $\neg p$, we have that $p \mid\!\sim \neg f$ is exceptional. It is the only exceptional formula therefore it moves to a different level to the other defeasible implications. This gives us the following ranking, shown in the table, which is equivalent to $(R_0, ..., R_{n-1}, R_\infty, n)$. The algorithm to build the ranking using exceptionality is also formally presented on the next page.

∞	$p \rightarrow b, r \rightarrow b$
1	$p \rightarrow \neg f$
0	$b \rightarrow f$,$b \rightarrow w$

Now we will go through the following steps if we want to check if a defeasible formula, let's say $r \mid\!\sim w$, is in the rational closure. We must then check if $\neg r$ is derivable from the classical version of the rankings. $\neg r$ is not derivable in this example so we can continue. If $\neg r$ was derivable then we would remove the top layer from the ranking and continue this process until we no longer derive $\neg r$. Once that is complete, then we can now check if $r \to w$ can be classically entailed from the statements that remain in the above ranking. In this case, they do, so we derive $r \mid\!\sim w$.

Algorithm 1: Ranking

```
Input: $\mathcal{KB}$
Output: An ordered tuple $(R_0, ..., R_{n-1}, R_\infty, n)$
i := 0;
$\mathcal{E}_0 := \overrightarrow{\mathcal{KB}}$;
while $\mathcal{E}_{i-1} \neq \mathcal{E}_i$ do
    $\mathcal{E}_{i+1} := \{\alpha \to \beta \in \mathcal{E}_i \mid \mathcal{E}_i \models \neg\alpha\}$;
    $R_i := \mathcal{E}_i \setminus \mathcal{E}_{i+1}$;
    $i := i + 1$;
$R_\infty := \mathcal{E}_{i-1}$;
if $\mathcal{E}_{i-1} = \emptyset$ then
    $n := i - 1$;
else
    $n := i$;
return $(R_0, ..., R_{n-1}, R_\infty, n)$
```

If we were to ask whether penguins have wings in this example, we would see that rational closure cannot derive $p \mid\!\sim w$. For the case with $p \mid\!\sim w$, $\neg p$ is true when first checked on the above classical ranking. Therefore the least exceptional layer, $\{b \to f, b \to w\}$, must be removed and rids the model of the statement $b \to w$. Without this statement we can only tell that penguins are birds but have no indication on whether birds have wings and there is therefore no way to derive that penguins have wings. This is because rational closure employs a prototypical reading. Using rational closure we cannot conclude that penguins have wings because it does not represent a typical bird and therefore cannot inherit any properties of the group it is a subclass of (penguin is a subclass of bird) [4]. Rational closure is said to be a conservative form of entailment since it is unable to derive some intuitive conclusions [1]. Now to formally present the rational closure algorithm which incorporates the ranking algorithm.

```
Algorithm 2: RationalClosure
  Input: $\mathcal{KB}$ and $\alpha \mid\!\sim \beta$
  Output: true, if $\mathcal{KB} \mid\!\approx \alpha \mid\!\sim \beta$, and false otherwise
  $(R_0, ..., R_{n-1}, R_\infty, n) :=$ Ranking($\mathcal{KB}$);
  $i := 0$;
  $R := \bigcup_{i=0}^{j<n} R_j$;
  while $R_\infty \cup R \models \neg\alpha$ and $R \neq \emptyset$ do
      $R := R \setminus R_i$;
      $i := i + 1$;
  return $R_\infty \cup R \models \alpha \rightarrow \beta$;
```

3.6 Lexicographic Closure

Lehmann detailed another form of entailment for defeasible reasoning called lexicographic closure [14]. The following are some of the properties that, along with properties that rational closure supports, guided the construction of lexicographic closure as stated by Lehmann [14].
Presumption of typicality tells us that if we have $\alpha \mid\!\sim \beta$ then we can either accept $\alpha \wedge \gamma \mid\!\sim \beta$ or $\alpha \mid\!\sim \neg\gamma$. This property tells that we only accept $\alpha \mid\!\sim \neg\gamma$ when we have concrete evidence to do so, otherwise, we presume the former. This property essentially tells us to assume the aforementioned property of monotonicity unless we are explicitly given information to the contrary. *Presumption of independence* states that if we have two consequents that result in the presumption of typicality being unable to choose which one to accept then the presumption of independence informs which to accept. This property broadly says to presume typicality for every consequent unless there is a reason for the contrary. For the previous example if we have $\alpha \mid\!\sim \beta$ and $\alpha \mid\!\sim \neg\gamma$ then we cannot use the presumption of typicality to derive $\alpha \wedge \gamma \mid\!\sim \beta$ but the presumption of independence tells us that β and γ are independent unless otherwise stated so this supports the derivation of $\alpha \wedge \gamma \mid\!\sim \beta$. *Priority to typicality* tells us to prefer inferences from the presumption of typicality over derivations from the presumption of indepen-

dence. *Respect for specificity* is a property that is also satisfied by rational closure and is now mentioned to be thorough. For this property, any two conflicting statements can be resolved using a preference for the statement with the more specific antecedent. This selection between the statements would be based more on intuition than any formal guideline.

Now to revisit the Bird example from section 3.5. We were unable to derive $p \mid\sim w$ from $\{b \mid\sim f, b \mid\sim w, p \mid\sim \neg f, p \rightarrow b, r \rightarrow b\}$ using rational closure but we show that we can derive $p \mid\sim w$ using lexicographic closure. The following are the summarised steps of a lexicographic closure algorithm for propositional logic by Casini et al., which was generalised to be implemented on description logics [4]. We will make use of the Bird example to illustrate the process more clearly. Firstly, we will separate the knowledge base into $\mathcal{A} = \{p \rightarrow b, r \rightarrow b\}$ and $\mathcal{B} = \{b \mid\sim f, b \mid\sim w, p \mid\sim \neg f\}$, which are the classical and defeasible parts of the knowledge base respectively. We use $\mathcal{A}$ and $\mathcal{B}$ along with the entire knowledge base, $\mathcal{KB}$, in various parts of the algorithm. When we refer to the premises, we are referring to the antecedent of the defeasible statement which we query. So if we were to query whether $p \mid\sim w$ is derived by lexicographic closure then p would be the premise.

Step 1 We create a set of materialisations of the statements in the knowledge base. This set of materialisations will be $\overrightarrow{\mathcal{KB}}$. We must then check the consistency of the knowledge base using the materialisations set. If a knowledge base is inconsistent then anything will follow from it and which is not desired when reasoning. A knowledge base $\mathcal{KB}$ is then inconsistent if and only if $\overrightarrow{\mathcal{KB}} \models \bot$. So $\overrightarrow{\mathcal{KB}}$ in the Bird example is $\{p \rightarrow b, r \rightarrow b, b \rightarrow f, b \rightarrow w, p \rightarrow \neg f\}$. This set is consistent, $\overrightarrow{\mathcal{KB}} \not\models \bot$, therefore we can continue with the algorithm. This is the case in this example despite the presence of the three statements $p \rightarrow b$, $b \rightarrow f$ and $p \rightarrow \neg f$ because we can have a situation where there no penguins and thus have no conflicts.

Step 2 We then give each statement in $\mathcal{B}$ a rank based on their exceptionality, which is similar to what done was in section 3.5. The ranks of the example's statements are as follows: $\{r(b \mid\sim f) = 0, r(b \mid\sim w) = 0, r(p \mid\sim \neg f) = 1\}$. The ranks for the statements in $\mathcal{A}$ will be infinite and for the example we have that $r(p \rightarrow b) = \infty$ and $r(r \rightarrow b) = \infty$.

Step 3 We define the set $\tilde{\mathcal{B}}$ to be $\{\alpha \mid\sim \beta \in \mathcal{B} \mid r(\alpha \mid\sim \beta) < \infty\}$. So $\tilde{\mathcal{B}}$ will be all the defeasible implications in $\mathcal{B}'$ with a rank less than infinity. The rank of $\tilde{\mathcal{B}}$, denoted by $r(\tilde{\mathcal{B}})$, will be the highest rank among the defeasible

implications. The example's $\tilde{\mathcal{B}}$ will be $\{b \mid\!\sim f, b \mid\!\sim w, p \mid\!\sim \neg f\}$ and $r(\tilde{\mathcal{B}}) = 1$..
Step 4 We will now define $\mathcal{T}$, which will be the set of the most preferred subsets of $\mathcal{X}$, where $\mathcal{X} = \{\alpha \rightarrow \beta \mid \alpha \mid\!\sim \beta \in \tilde{\mathcal{B}}\}$, which satisfy both our premises and $\mathcal{A}$. Take k to be the rank of $\tilde{\mathcal{B}}$ and define the subset $\mathcal{X}^i$ as the subset of conditionals in $\mathcal{X}$ which have rank i. So we can now give every subset $\mathcal{D}$ of $\mathcal{X}$ a sequence of natural numbers with each number representing the number of statements in that subset that have a certain rank. For the sequence, $<n_0, ..., n_k>_\mathcal{D}$, the numbers are generally defined as $n_i = \mid \mathcal{D} \cap \mathcal{X}^{k-i} \mid$. It is with these sequences of numbers that we rank the subsets, where the number of statements they satisfy is important as well as the exceptionality of the statements they satisfy. We can say that a subset $\mathcal{D}$ is preferred to $\mathcal{E}$ iff $<n_0, ..., n_k>_\mathcal{D} > <n_0, ..., n_k>_\mathcal{E}$. And we say that $<n_0, ..., n_k> \geq <m_0, ..., m_k>$ iff (i) for every i, such that $0 \leq i \leq k$, $n_i \geq m_i$ or (ii) if $n_i < m_i$, then there is a j such that $j < i$ and $n_j > m_j$.
So now to define the set $\mathcal{T}$ for our example. We list the subsets of $\mathcal{X}$ along with their natural number sequence where $r(\tilde{\mathcal{B}}) = 1$.

- $\{b \rightarrow f, b \rightarrow w, p \rightarrow \neg f\} \mid < 1, 2 >$
- $\{b \rightarrow f, b \rightarrow w\} \mid < 0, 2 >$
- $\{b \rightarrow f, p \rightarrow \neg f\} \mid < 1, 1 >$
- $\{b \rightarrow w, p \rightarrow \neg f\} \mid < 1, 1 >$
- $\{b \rightarrow f\} \mid < 0, 1 >$
- $\{b \rightarrow w\} \mid < 0, 1 >$
- $\{p \rightarrow \neg f\} \mid < 1, 0 >$

So for premise p and $\mathcal{A} = \{p \rightarrow b, r \rightarrow b\}$, $\mathcal{T}$ will be $\{\{b \rightarrow w, p \rightarrow \neg f\}\}$.
Step 5 Finally, given a set of premises, p, we say $p \mid\!\sim w$ is in the lexicographic closure if $p \cup \mathcal{A} \cup \mathcal{D} \models w$ for every $\mathcal{D} \in \mathcal{T}$. So if we take p to be the premise, we can see that $p \mid\!\sim w$ is in the lexicographic closure and this can be denoted by $p \mid\!\sim^{lc}_{\mathcal{KB}} w$.

$$\{p\} \cup \{b \rightarrow w, p \rightarrow \neg f\} \cup \{p \rightarrow b, r \rightarrow b\} \models w$$

Note that the statements derived by rational closure will always be in the lexicographic closure, therefore, lexicographic closure can be thought of as the more venturous of the two entailment methods [1, 4, 14]. Lexicographic closure offers a different interpretation to rational closure for defeasible scenarios. As previously mentioned, with a prototypical reading, the statement $b \mid\!\sim w$ tells us that typical birds have wings and this is what we get with rational closure. This means that if we have an atypical bird then we cannot conclude that the bird has wings. An example being penguins, atypical birds in that they do not fly, which the rational closure algorithm states do not have wings. The reading with lexicographic closure is a presumptive one. The presumptive reading of $b \mid\!\sim w$ presumes that birds have wings unless we are provided evidence to the contrary. This is the case even if we have an atypical bird such as a penguin. And we have seen already that the lexicographic closure algorithm gives us that penguins have wings. The lexicographic closure algorithm is presented formally and also makes use of the ranking algorithm presented in section 3.5.

```
Algorithm 3: LexicographicClosure
  Input: $\mathcal{KB}$ and $\alpha \mid\!\sim \beta$
  Output: true, if $\alpha \mid\!\sim^{lc}_{\mathcal{KB}} \beta$, and false otherwise
if $\overrightarrow{\mathcal{KB}} \models \bot$ then
    return false
$\mathcal{A} = \{\alpha \rightarrow \beta \in \mathcal{KB}\}$;
$\mathcal{B} = \{\alpha \mid\!\sim \beta \in \mathcal{KB}\}$;
$(R_0, ..., R_{n-1}, R_\infty, n)$ := Ranking($\mathcal{KB}$);
$\tilde{\mathcal{B}} = \{\alpha \mid\!\sim \beta \in \mathcal{B} \mid r(\alpha \mid\!\sim \beta) < \infty\}$;
$\mathcal{X} = \{\alpha \rightarrow \beta \mid \alpha \mid\!\sim \beta \in \tilde{\mathcal{B}}\}$;
$r(\mathcal{B}) = max\{\{r(a \mid\!\sim \beta) \mid \alpha \mid\!\sim \beta \in \mathcal{B}\}\}$;
$\forall \mathcal{D} \in \mathcal{X}$ assign $< n_0, ..., n_k >_{\mathcal{D}}$, where $n_i =\mid \mathcal{D} \cap \mathcal{X}^{k-i} \mid$ with $k = r(\tilde{\mathcal{B}})$;
$\geq_{\mathcal{X}}$ iff $(i)$ $\forall i$, s.t. $0 \leq i \leq k, n_i \geq m_i$ or $(ii)$ if $n_i < m_i$, then there is a $j$
  such that $j < i$ and $n_j > m_j$;
$\mathcal{T} = \{\mathcal{D} \subseteq \mathcal{X} \mid \forall \mathcal{E} \subseteq \mathcal{X}, \mathcal{D} \geq_{\mathcal{X}} \mathcal{E}\}$;
return $\alpha \cup \mathcal{A} \cup \mathcal{D} \models \beta, \forall \mathcal{D} \in \mathcal{T}$
```

Chapter 4

Deontic Logic

This chapter will formally present deontic logic in general and the specific logic system we will investigate. Deontic Logic is a field of logic which formalises normative concepts. These concepts include obligation ("what is an individual's duty", "what an individual ought to do"), permission ("what an individual may do") as well as other related concepts such as prohibition ("what an individual is forbidden from doing") [11, 22, 23]. The system we will work with is the traditional Dyadic Standard Deontic Logic (DSDL) approach [21, 22, 23], although there are alternative approaches to deontic logic such as input/output logic [11, 17]. The reason we opted for the more traditional approach was that it has semantics based on valuations, similar to that of the other logic systems we deal with [21, 23]. This ensures a connection between the logic systems which we can leverage to get results during analysis. With DSDL extending the standard deontic system aptly named Standard Deontic Logic (SDL) [21, 22, 23], we will briefly present SDL. We represent the previously mentioned normative concepts in SDL using the notation: $\bigcirc p$ to say that "p is obligatory", $\mathrm{P}p$ to say that "p is permitted" and $\mathrm{F}p$ to say that "p is forbidden". Dyadic Standard Deontic Logic (DSDL) is then a form of deontic logic which better handles conditional obligations such as "if p is true then you must do q". We can represent such statements using the "$|$" notation which is usually seen in conditional probability. The example statement, "if p is true then you must do q", would be represented by $\bigcirc(q \mid p)$ in DSDL. Since many legal statements are of the conditional form, we use DSDL as the logic when we are dealing in the deontic environment instead of SDL.

4.1 Language

Definition 4.1. *Given a set of propositional letters Φ, the language of DSDL can be formed with the following operator added to the propositional logic language [22, 23]: $\bigcirc$ is the operator added which represents obligation as previously mentioned. This operator can be used similarly to the negation operator $\neg$ in that we can place it in front of any propositional formula and we can also apply it in a nested fashion such as in the following example DSDL formula $\bigcirc(p \mid q \wedge \bigcirc(r \mid p))$ [22].*

The notion of permission being related to obligation by $\mathrm{P}p = \neg\bigcirc\neg p$ and that of prohibition being similarly related by $\mathrm{F}p = \bigcirc\neg p$. We can write non-conditional obligations as conditional ones by $\bigcirc p = \bigcirc(p| \top)$ [23]. We can now formally define the preference-based semantics for DSDL which were presented with similar formal definitions by Parent et al. [22] and Pigozzi et al. [23].

4.2 Semantics

These semantics are similar to the ranked-interpretation semantics of KLM defeasible reasoning.

Definition 4.2. *We have preference models defined as $M = (V, \leq)$ where $V \subseteq W$, with W being a non-empty set of possible valuations. $\leq$ is not only a binary relation over V but a total preorder as it is reflexive, transitive and connected. Connectivity states that $\forall\, v,\, w \in V$, then either $v \leq w$ or $w \leq v$ or both. The operator $\models$ represents the satisfaction of a formula. Let's say we have a valuation $s \in V$ as well as propositional formulas α and β. Satisfaction is done in the classical way of propositional logic with the addition of the following [22]:*

- *$s \models \bigcirc(\beta \mid \alpha)$ iff $\forall\, s' \in V$, if $s' \in \{s \in [\![\alpha]\!] : s \leq t, \forall\, t \in [\![\alpha]\!]\}$, then $s' \models \beta$. Here $[\![\alpha]\!] = \{s \in V : s \models \alpha\}$. $s < s'$ means that $s \leq s'$ and $s' \not\leq s$. This means that given α being true then only if the "best" or "most typical" valuations that satisfy α also satisfy β then we can derive $\bigcirc(\beta \mid \alpha)$ and therefore say β is obligatory given α. This is similar to the model semantics of the KLM-style defeasible implications as stated in chapter 3. Note that there are other possible definitions of "best"*

> *valuations, such as the one provided by Hansson [10]. For the KLM approach, we say that a defeasible implication $\alpha \mid\sim \beta$ is true if the minimal or 'best' valuations that satisfy α also satisfy β.*

Note that we will not allow for duplicate valuations, i.e for all $v_1, v_2 \in W$ such that $v_1 \neq v_2$, then there is a propositional formula, let's say α, such that $v_1 \models \alpha$ and $v_2 \not\models \alpha$. This is a definition similar to that used by Van der Torre when discussing the two-phase deontic logic [27]. This is necessary to prevent a scenario, let's say we have valuations v_1, v_2 and v_3 with $v_! < v_2 < v_3$, where we wish to have v_2 preferred to v_3 but then have a v_1 which is a duplicate of v_3 and is preferred to both v_2 and v_3.

4.3 Deontic Properties

The following outlines some properties that commonly occur in the deontic logic literature [8, 21, 23, 27]. We chose these properties because they were seen as being important or at least relevant when assessing the usefulness of a deontic logic system. Thus they should be thought of as properties that an ideal deontic logic system would have. We give justification along with each property to show why one may want such a property in a deontic logic system. We present these justifications primarily in the form of examples that we intend to demonstrate the intuition behind selecting the property. Note that this is not a full list of properties that are desirable for a deontic logic nor are they necessary properties for a reasonable deontic system. In that regard, this list of properties is different to that for the rationality of the entailment relation, $\models\!\approx$, in chapter 3 as the non-satisfaction of any these properties by a deontic system, does not necessarily mean that the deontic system will be less effective whereas the rationality of an entailment relation relies on the satisfaction of those properties in chapter 3. These deontic properties are those we deemed desirable, but interpretations of each property's appropriateness within a deontic setting may differ from ours. Despite this difference, we will re-use the entailment relation operator, $\models\!\approx$, to denote deontic derivations to form a connection between DSDL and the KLM approach. $\mathcal{KB}$ in these derivations will refer to a deontic knowledge base which is a knowledge base which can contain deontic formulas in addition to those of propositional logic. As usual, $\models$ is used to represent classical entailment.

Restricted Strengthening of the Antecedent

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\beta|\alpha)}{\mathcal{KB} \mid\!\approx \bigcirc(\beta|\alpha \wedge \gamma)}$$

Let's say we have the obligation to do β when α is true. It is intuitive that a more specific version of α being true would still make β obligatory since α will remain to be true if a more specific version of it is true. Here $\alpha \wedge \gamma$ is a more specific version of α. If we have that *"if you are a student then you must have a student card"* then we should be able to derive that *"if you are a student and a male then you must have a student card"*. Note that the restricted version of the property that we refer to requires the formula $\alpha \wedge \gamma$, of the obligation $\bigcirc(\beta \mid \gamma \wedge \alpha)$ that we derive, to be consistent. The property will be referred to as Restricted Strengthening of the Antecedent during this dissertation. In general, we do not desire Strengthening of the Antecedent since we wish for obligations to be defeasible [22]. We instead check whether there are certain conditions where we can apply it in a similar manner to the Cautious Monotonicity and Rational Monotonicity properties from section 3.4.

Factual Detachment

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\beta|\alpha), \models \alpha}{\mathcal{KB} \mid\!\approx \bigcirc\beta}$$

If we have an obligation to do a task β when α is satisfied, once we have that α has occurred then it is intuitive that we are now obligated to do β. If we have that *"if you are a student then you must have a student card"* and are presented with the fact that *"you are a student"* then we should be able to derive that *"you must have a student card"*.

Deontic Detachment

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\beta|\alpha), \mathcal{KB} \mid\!\approx \bigcirc\alpha}{\mathcal{KB} \mid\!\approx \bigcirc\beta}$$

If we have an obligation to do a task β when α is satisfied and we have that α is obligated to be true then it is intuitive that we are now obligated to do β. If we have that *"if you are a student then you must have a student card"* and we also have that *"you have to be a student"* then we should be able to derive that *"you must have a student card"*.

Transitivity

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\gamma|\beta), \mathcal{KB} \mid\!\approx \bigcirc(\beta|\alpha)}{\mathcal{KB} \mid\!\approx \bigcirc(\gamma|\alpha)}$$

Let's say we have an obligation to do a task β when α is satisfied. Then we also have an obligation to do γ when β is satisfied. By combining these two obligations, it is intuitive that we are now obligated to do γ if we have α. If we have that *"if you are a lecturer then you must be a staff member"* and we also have that *"if you are a staff member then you must have a staff number"* then we should be able to derive *"if you are a lecturer then you must have a staff number"*.

Disjunction

$$\frac{\mathcal{KB}\mid\!\approx\bigcirc(\beta|\gamma),\mathcal{KB}\mid\!\approx\bigcirc(\beta|\alpha)}{\mathcal{KB}\mid\!\approx\bigcirc(\beta|\alpha\vee\gamma)}$$

Let's say we have an obligation to do a task β when α is satisfied. Then we also have an obligation to do the same β when γ is satisfied. By combining these two obligations it is intuitive that we are now obligated to do β when we have either α or γ. If we have that *"if you are a student trying to enter the library then you must have a student card"* and we also have *"if you are a student trying to enter the computer lab then you must have a student card"* then we should be able to derive *"if you are a student trying to enter the library or a student trying to enter the computer lab then you must have a student card"*.

Conjunction

$$\frac{\mathcal{KB}\mid\!\approx\bigcirc(\beta|\alpha),\mathcal{KB}\mid\!\approx\bigcirc(\gamma|\alpha)}{\mathcal{KB}\mid\!\approx\bigcirc(\beta\wedge\gamma|\alpha)}$$

Let's say we have an obligation to do a task β when α is satisfied and we also have an obligation to do γ when α is satisfied. By combining these two obligations it is intuitive that we are now obligated to do both β and γ if when we have α. If we have, for example, that *"if you are a student then you must have a student card"* and also have *"if you are a student then your fees must be paid in full"* then we would want it to be the case that *"if you are a student then you must have a student card and must have your fees paid for"*.

Weakening

$$\frac{\mathcal{KB}\mid\!\approx\bigcirc(\beta\wedge\gamma|\alpha)}{\mathcal{KB}\mid\!\approx\bigcirc(\beta|\alpha)}$$

Note that $\beta \wedge \gamma$ implies β. Let's say that we have the obligation to do both γ and β when α is true. It is intuitive that we can derive an obligation to do only one of β or γ when α is satisfied. If we have, for example, that *"if you are a student then you must have a student card and have your fees paid for"* then we should be able to derive *"if you are a student then you must have a student card"*.

Ought Implies Can

$$\neg \bigcirc (\alpha \wedge \neg\alpha)$$

This property could also be represented as $\neg \bigcirc \bot$ as the conjunction of conflicting tasks, $\alpha \wedge \neg\alpha$, will be a logical contradiction and can therefore be represented by $\bot$. The property states that if there is an obligation then it should be possible to fulfil it. So it should not be obligatory to do contradictory tasks such as α and $\neg\alpha$. Now this brings the dilemma of how to detect violations, as these will be situations where contradictions involving obligations occur. If we cannot allow such a conflict, then there is a need to find a method for violation detection [21, 23].

Distribution

$$\text{If} \models \alpha \to \beta, \text{ then } \mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\alpha \to \bigcirc\beta$$

Let's say we know that $\alpha \to \beta$ is a tautology and then find out that α being true is obligatory. It is then intuitive to require that β is also obligatory. If we have that *"students write exams"* and have that *"you have to be a student"* then we should be able to derive that *"you have to write exams"*.

Deontic Explosion Principle

Finally, we present a principle which we desire of a deontic system and that principle is the underivability of deontic explosion. Let's say we have a knowledge base with $\bigcirc\alpha$ and $\bigcirc\neg\alpha$ as well as the properties of Weakening and Conjunction. The following is a possible derivation of deontic explosion.

$$\cfrac{\cfrac{\cfrac{\mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\alpha, \mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\neg\alpha}{\mathcal{KB} \mathrel{|\!\!\approx} \bigcirc(\alpha \vee \beta), \mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\neg\alpha}}{\mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\beta \wedge \neg\alpha}}{\mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\beta}$$

Deontic explosion is where any obligation can be derived once an inconsistency occurs such as when there is both $\bigcirc\alpha$ and $\bigcirc\neg\alpha$ [8, 21]. This occurs from the combination of Conjunction and Weakening [21]. This an undesirable occurrence since a conflict between two obligations in the actual world does not allow for any other act to be obligatory. This is an issue that will not affect an ideal system.

4.4 Paradoxes

A deontic paradox is a set of deontic conditionals that give a counter-intuitive result even though the conditionals themselves are consistent and intuitive. We chose these paradoxes as they occur frequently in the deontic logic literature [8, 21, 23, 27]. One of the reasons that the paradoxes are primarily used to analyse the logic systems is that the paradoxes are of similar structure to many other deontic examples [27]. They also provide difficulty to the logics that the straightforward examples would not, as they posed a challenge for deontic logic researchers for many years [20, 21, 27]. The paradoxes will be presented and the derivations of their issues detailed. Some of them have multiple issues presented in order to show there being multiple paths to problems within the paradoxes. For obligations $\bigcirc(\beta_1 \mid \alpha_1)$ and $\bigcirc(\beta_2 \mid \alpha_2)$, we say that the second obligation is a contrary-to-duty obligation of the first if its antecedent α_2 is contradictory to the consequent of the first, β_1 [8, 21, 23, 27]. Intuitively, this means an obligation that informs us what must be the case when something forbidden has been done [26]. We can also say that the second obligation is an according-to-duty obligation of the first if its antecedent α_2 logically implies the consequent of the first β_1. This tells us what to do when an obligation has been fulfilled [21, 23, 27].

The following figures show derivations of obligations by using an arrow with a subscript containing the abbreviation of the property which was used for the derivation. $\bigcirc(\beta \mid \alpha) \rightarrow_W \bigcirc(\gamma \mid \alpha)$ means that Weakening was used to go from $\bigcirc(\beta \mid \alpha)$ to $\bigcirc(\gamma \mid \alpha)$. Weakening would be an applicable property in this example if we knew $\beta \rightarrow \gamma$ to always be true. The derivations that involve more than one obligation as the premise have the obligations displayed between braces and separated by a comma. The derivation $\{\bigcirc(\gamma \mid \alpha), \bigcirc(\beta \mid \alpha)\} \rightarrow_{Conj} \bigcirc(\gamma \wedge \beta \mid \alpha)$ means that the Conjunction property was used on the obligations $\bigcirc(\gamma \mid \alpha)$ and $\bigcirc(\beta \mid \alpha)$ to derive $\bigcirc(\gamma \wedge \beta \mid \alpha)$.

4.4.1 Forrester's paradox

This paradox comprises three statements: two obligations and a fact. *"You must not kill anybody"*, *"If you kill someone then you must kill them gently"* and *"You killed someone "*. With these we also have the background knowledge that *"Killing gently implies killing"*. The paradox's statements can be represented by the following set of deontic statements.

$$\{\bigcirc\neg k, \bigcirc(g \mid k), k\}$$

Restricted Strengthening of the Antecedent, Weakening and Conjunction

1.	$\bigcirc\neg k \rightarrow_W \bigcirc\neg g$
2.	$\bigcirc\neg g \rightarrow_{RSA} \bigcirc(\neg g \mid k)$
3.	$\{\bigcirc(\neg g \mid k), \bigcirc(g \mid k)\} \rightarrow_{Conj} \bigcirc(\neg g \wedge g \mid k)$

We represent the background knowledge with the propositional formula $g \rightarrow k$ and we assumed it to hold throughout. When we apply Weakening we can derive the obligation *“You must not kill gently”* from the first obligation, *“You must not kill anybody”*. This is sensible since killing gently is still an act of killing which we want to be forbidden. We proceed to use Restricted Strengthening of the Antecedent and the fact *“You killed someone”* to move from the non-conditional obligation *“You must not kill gently”* to the conditional obligation *“If you kill then you must not kill gently”*. This derivation is an issue with the paradox as it is counter-intuitive for an obligation to be the premise from which its own contrary-to-duty obligation is derived [21, 23]. Finally, using Conjunction we can derive a contradiction from the obligations *“If you kill then you must not kill gently”* and *“If you kill then you must kill gently”*.

Factual Detachment and Conjunction

1.	$\{\bigcirc(g \mid k),\ k\} \rightarrow_{FD} \bigcirc g$
2.	$\{\bigcirc\neg k,\ \bigcirc g\} \rightarrow_{Conj} \bigcirc(\neg k \wedge g)$

The rule of Factual Detachment gives us the non-conditional obligation *“You must kill gently”* from the fact *“You killed someone”* and the conditional obligation *“If you kill then you must kill gently”*. Applying the property Conjunction to the conditional obligation *“You must kill gently”* and the non-conditional obligation *“You must not kill anybody”* produces the following: *“You must not kill and you must kill gently”*. This is a contradiction thus the derivation of this conjunction is undesirable [21, 23].

4.4.2 Chisholm’s paradox

This paradox comprises four statements: three obligations and a fact. *“Jones must go assist his neighbour”*, *“If Jones goes to assist his neighbour then he* *must tell them that he is coming"*, *"If Jones does not go to assist his neighbour then he must not tell them that he is coming"* and *"Jones does not go to assist his neighbour"*. The paradox's statements can be represented by the following set of deontic statements.

$$\{\bigcirc a, \bigcirc(t \mid a), \bigcirc(\neg t \mid \neg a), \neg a\}$$

Deontic Detachment, Factual Detachment and Conjunction

1.	$\{\bigcirc(t \mid a),\ \bigcirc a\} \rightarrow_{DD} \bigcirc t$
2.	$\{\bigcirc(\neg t \mid \neg a),\ \neg a\} \rightarrow_{FD} \bigcirc\neg t$
3.	$\{\bigcirc t, \bigcirc\neg t\} \rightarrow_{Conj} \bigcirc t \wedge \neg t$

The non-conditional obligation *"Jones must go assist his neighbour"*, in combination with the conditional obligation *"If Jones goes to assist his neighbour then he must tell them he is coming"* gives us, via Deontic Detachment, the obligation that *"Jones must tell his neighbour he is coming"*. Factual Detachment also gives the non-conditional obligation *"Jones must not tell his neighbour he is coming"* since we have the conditional obligation *"If Jones does not go to assist his neighbour then he must not tell them he is coming"* and the fact that John does not go help his neighbour. There is then a conflict as we can use Conjunction to generate *"Jones must tell his neighbour he is coming and Jones must not tell his neighbour he is coming"*. It does not seem reasonable to reject Conjunction in this case so the problem seems to be between the use Factual Detachment and Deontic Detachment.

In the literature, there seems to have been a divide between supporters of Factual Detachment and those of Deontic Detachment. Those in support of Factual Detachment often reject Deontic Detachment on the basis that obligations such as the second in Chisholm's paradox, *"If Jones goes to assist his neighbour then he must tell them he is coming"*, only tell us what to do in ideal situations and do not give us actual obligations to act upon once a primary obligation has been violated [11]. On the other hand, Hilpinen et al. [11] detail an obligation that those in favour of Deontic Detachment often point to. This obligation is *"If Jones will kill his rich aunt now (for the inheritance), then he ought to shoot her to death"* and let's say we also know for a fact that he will indeed kill her although he does not have to. Since it is a fact that he will kill her then he is obligated to shoot her. This then turns the obligation not to kill her into only an ideal obligation which does not seem reasonable. This seems to say that Factual Detachment allows for a person to act immorally in order to generate an actual obligation that they can fulfil, which is less immoral. We say this seems unreasonable because in the end, their obligation would still be to act immorally [11]. Another example Deontic Detachment supporters might cite is the pragmatic oddity of Prakken and Sergot [11, 27], which is analogous to the aforementioned Forrester's paradox. Ideally, we wish to sacrifice neither property when we conduct the deontic analysis in later chapters.

Transitivity and Restricted Strengthening of the Antecedent

1.	$\{\bigcirc(t \mid a), \bigcirc a\} \rightarrow_{Trans} \bigcirc t$
2.	$\bigcirc t \rightarrow_{RSA} \bigcirc(t \mid \neg a)$
3.	$\{\bigcirc(t \mid \neg a), \bigcirc(\neg t \mid \neg a)\} \rightarrow_{Conj} \bigcirc(t \wedge \neg t \mid \neg a)$

Similarly to the derivations from Deontic Detachment, the application of the Transitivity property can take us from the obligations *"Jones must go assist his neighbour"* and *"If Jones goes to assist his neighbour then he must tell them that he is coming"* to the non-conditional obligation *"Jones must tell his neighbour he is coming"*. Restricted Strengthening of the Antecedent applied to that obligation then gives us the counter-intuitive *"If Jones does not go assist his neighbour then he must tell them he is coming"*. Since we also have the conditional obligation *"If Jones does not go to assist his neighbour then he must not tell them that he is coming"*, this gives us a contradiction.

4.4.3 Fence paradox

This paradox comprises four statements: three obligations and a fact. *"There must not be a fence"*, *"If there is a fence then it must be a white fence"*, *"If there is a dog then there must be a white fence"* and *"There is a fence"*. The paradox's statements can be represented by the following set of deontic statements.

$$\{\bigcirc\neg f, \bigcirc(f \wedge w \mid f), \bigcirc(f \wedge w \mid d), f\}$$

The main point of interest regarding this paradox is concerned with the distinguishing between a violation and exception whenever there is a fence. If there is a fence and a dog is present then this represents an exception and the first obligation intuitively is not violated but is rather overridden. Whereas the presence of a fence without a dog should be considered a violation. So

$\bigcirc(\neg f \mid f)$ should be derivable but not $\bigcirc(\neg f \mid f \wedge d)$ which is a problem if there is a form of the property Strengthening the Antecedent. This property allows for that derivation of $\bigcirc(\neg f \mid f \wedge d)$ when a dog is present [21, 23, 27].

4.4.4 Trump/Kim paradox

This paradox is an alphabetic variant of the Reykjavik scenario [27]. This version comprises four obligations. *"Trump must not be told the secret"*, *"Kim must not be told the secret"*, *"If you tell Trump then you must tell Kim"* and *"If you tell Kim then you must tell Trump"*. The obligations can be represented by the following set of deontic statements.

$$\{\bigcirc\neg t, \bigcirc\neg k, \bigcirc(k \mid t), \bigcirc(t \mid k)\}$$

Restricted Strengthening of the Antecedent, Weakening and Conjunction

1.	$\{\bigcirc\neg t, \bigcirc\neg k\} \rightarrow_{Conj} \bigcirc(\neg t \wedge \neg k)$
2.	$\bigcirc(\neg t \wedge \neg k) \rightarrow_{W} \bigcirc\neg(t \wedge k)$
3.	$\bigcirc\neg(t \wedge k) \rightarrow_{RSA} \bigcirc(\neg(t \wedge k) \mid t)$
4.	$\{\bigcirc(\neg(t \wedge k) \mid t), \bigcirc(t \wedge k \mid t\} \rightarrow_{Conj} \bigcirc(\neg(t \wedge k) \wedge t \wedge k \mid t)$

This paradox is another example where the combination of Weakening and Restricted Strengthening of the Antecedent causes problems. The intuitive obligation *"Do not tell Trump and do not tell Kim"* using conjunction can bring forth the derivation of *"Do not tell both Trump and Kim"* using Weakening. This seems reasonable as a non-conditional obligation when in combination with the obligations that we should not tell either of them. But from Restricted Strengthening of the Antecedent we are able to derive *"If you tell Trump then do not tell both Trump and Kim"* which is a problem because from the set of premises, we want it to be the case that when we tell one, we must tell the other. This gives us a conflict. Another issue with this paradox, besides the one concerning an application of properties, is in how to interpret the obligations [27]. In much the same fashion as other paradoxes, one must decide whether the statements $\{\bigcirc(k \mid t), \bigcirc(t \mid k)\}$ should be treated as exceptions to the first two and therefore when we have $t \wedge k$ to be true, there is no violation. Or if we want the derivation of $\bigcirc(\neg k \mid t \wedge k)$ and $\bigcirc(\neg t \mid t \wedge k)$, which explicitly tell us there is a violation. Van der Torre states that the latter interpretation is the preferred one [27] and later in the dissertation, we see that the tools we use for analysing the paradoxes will guide which interpretation we aim for.

4.4.5 Van Fraassen's paradox

This paradox comprises two obligations. *"You must honour your father or honour your mother"* and *"You must not honour your mother"*. The obligations can be represented by the following set of deontic statements.

$$\{\bigcirc(f \vee m), \bigcirc\neg m\}$$

Weakening and Conjunction

1.	$\{\bigcirc(f \vee m),\ \bigcirc\neg m\} \rightarrow_{Conj} \bigcirc(f \wedge \neg m)$
2.	$\bigcirc(f \wedge \neg m) \rightarrow_{W} \bigcirc f$

This paradox illustrates the issue of deontic explosion when we have the properties of Conjunction and Weakening. This first model shows how the desired obligation *"You must honour your father"* can be derived using these properties. Combining *"You must honour your father or honour your mother"* and *"You must not honour your mother"* tells one to honour their father. This justifies the desire to have these properties in a deontic system as the result is intuitive. But the following derivations, shows that once there is a conflict, the properties allow one to derive any other obligation. If we were to add the obligation *"You must honour your mother"* to the above paradox then this will be analogous to the following example generalised using formulas α and β. This further demonstrates the issue with deontic explosion.

1.	$\bigcirc\alpha \rightarrow_{W} \bigcirc(\alpha \vee \beta)$
2.	$\{\bigcirc(\alpha \vee \beta),\ \bigcirc\neg\alpha\} \rightarrow_{Conj} \bigcirc(\neg\alpha \wedge \beta)$
3.	$\bigcirc(\neg\alpha \wedge \beta) \rightarrow_{W} \bigcirc\beta$

There is another similarly structured example presented in by Van der Torre [27] called the Apples-and-Pears example. This example shows that there is also a problem with interpretation when Restricted Strengthening of the Antecedent is used on the example's set of obligations. The example has the set of obligations, $\{\bigcirc(a \vee p), \bigcirc\neg a\}$ which represent the statements, *"You must buy apples or buy pears"* and *"You must not buy apples"*. One could have a conditional interpretation where the derivation of $\bigcirc(p \mid a)$ is acceptable because we have $\bigcirc p$ informing us that pears must always be bought upon the condition $\top$. Whereas one can interpret the example contextually, so once we know that apples have been bought then there is no intuitive reason why pears must be bought. The obligation $\bigcirc a \vee p$ has already been fulfilled and the obligation $\bigcirc \neg a$ has already been violated. Therefore, we should not derive $\bigcirc (p \mid a)$ [27].

1.	$\{\bigcirc a \vee p, \bigcirc \neg a\} \rightarrow_{Conj} \bigcirc \neg a \wedge p$
2.	$\bigcirc \neg a \wedge p \rightarrow_{W} \bigcirc p$
3.	$\bigcirc p \rightarrow_{RSA} \bigcirc (p \mid a)$

Van Fraassen's paradox has been chosen to present an example that specifically deals with deontic explosion, while the Apples-and-Pears example presents a more intuitive reading of the obligations for the discussion of conditional and contextual interpretations.

Chapter 5

Propositional Typicality Logic

In this chapter, we present Propositional Typicality Logic (PTL), the logic system with which we explore the notion of typicality. PTL is formed by enriching classical propositional logic with an explicit operator to represent typicality [1].

5.1 Language

Definition 5.1. *Given a set of propositional letters* Φ*, the language of PTL, denoted by* $\mathcal{L}^{\bullet}$*, is formed with the bullet operator,* $\bullet$*, added to the propositional logic language [1]. This operator can be attached in front of any propositional formula. For example, for a propositional letter* p*, we could have* $\bullet p$ *in* $\mathcal{L}^{\bullet}$ *with its intuition being that it represents the most typical situations where* p *holds.*

We can use this operator similarly to the negation operator and placed in front of any propositional formula, and we can apply in a nested fashion such as in the following example PTL formula $\bullet\bullet p$. Note that this means that PTL is more expressive than KLM-style defeasible reasoning [1] and deontic logic as the bullet operator can be applied to both the antecedent and consequent side of a conditional. The following PTL statements illustrate the type of statements that can be represented with the PTL language. $\bullet p \rightarrow q$ reads as *"the most typical situations where* p *holds, imply the situations where* q *holds"*, $\bullet p \rightarrow \bullet\neg q$ reads as *"the most typical situations where* p *holds, imply the most typical situations where* q *does not hold"* and $(\bullet p \vee \bullet q) \rightarrow \bullet r$ reads as *"the most typical situations where* p *holds or the most typical situations* *where q holds, imply the most typical situations where r holds".* We now present the semantics for PTL, which utilises the familiar concept of ranked interpretations.

5.2 Semantics

Definition 5.2. *Let's say we have W being the set of possible valuations, ranked interpretations are pairs $< V, \leq >$, where $V \subseteq W$ and $\leq$ is a total preorder over V. Intuitively, the valuations pushed lower down the rankings are more typical than those that are higher up [1]. Note that $[\![\alpha]\!]^{\mathscr{R}}$ represents the set of valuations that satisfy a formula α for a given ranked interpretation $\mathscr{R}$ [1]. Satisfaction of a formula, let's says α, is done in the classical way of propositional logic with the addition of the following [1]:*

- *$v \models \bullet\alpha$ iff $v \models \alpha$ and there is not a $v' \leq v$ such that $v' \models \alpha$. So the valuations that satisfy $\bullet\alpha$ will be the minimal valuations that satisfy a. So $[\![\bullet\alpha]\!]^{\mathscr{R}} := min_{\leq}([\![\alpha]\!]^{\mathscr{R}})$ for a ranked interpretation $\mathscr{R}$.*

Note that the typicality $\bullet$-operator can be used to express any KLM-style conditional. That is, for every ranked interpretation $\mathscr{R}$ and every $\alpha, \beta \in \mathcal{L}$, $\mathscr{R} \models \alpha \mid\!\sim \beta$ if and only if $\mathscr{R} \models \bullet\alpha \rightarrow \beta$. There are $\mathcal{L}^{\bullet}$-sentences that cannot be expressed using KLM-style $\mid\!\sim$-statements on $\mathcal{L}$, so the converse does not hold [1]. Now we outline the first method of entailment we will use, which Booth et al. [1] proposed.

5.3 LM-entailment

The first form of entailment to be looked at is one that produces a single ranked model that is constructed to be the LM-minimum model for the knowledge base where LM refers to Lehmann and Magidor [1]. A sequence of ranked interpretations ($\mathscr{R}_0$,$\mathscr{R}_1$,$\mathscr{R}_2$,...), which is created during the algorithm, will be used to construct $\mathscr{R}^*_{\mathcal{KB}}$, which will be the model used for entailment. The algorithm will make use of ranks to construct $\mathscr{R}^*_{\mathcal{KB}}$. The ranks represent a level in the ranked interpretation, where the rank of a valuation u is less than the rank of v if and only if $u < v$, as defined in definition 5.2 [1]. The following brief explains some notation used during the LM-entailment algorithm. In this algorithm, we say $\mathscr{R}^1_S$ is the ranked interpretation obtained when any valuation not in S, where $S \subseteq V^{\mathscr{R}}$, has its rank increased by 1. Similarly, $\mathscr{R}_S^{\infty}$ is the ranked interpretation obtained from $\mathscr{R}$ by setting the rank of all valuations not in S to ∞ [1]. These would be those at the highest level of $\mathscr{R}^*_{\mathcal{KB}}$ and deemed to be atypical. Now to present the algorithm steps [1].

Step 1 Set the ranks of all valuations in the knowledge base to 0, define S_0 which is initially empty and have variable i equal to 1.
Step 2 Find the valuations which satisfy the knowledge base with respect to the current ranked interpretation $\mathscr{R}_0$ and put them into the set S_1.
Step 3 If S_i is equal to S_{i-1} then there has not been a change so set the rank of all the valuations that do not satisfy the knowledge base, with respect to $\mathscr{R}_i$, to ∞ and return the interpretation that remains.
Step 4 Otherwise create a new ranked interpretation $\mathscr{R}_i$, by increasing the rank of every valuation not in S_i by 1.
Step 5 Find the valuations which satisfy the knowledge base with respect to the current ranked interpretation $\mathscr{R}_i$ and put them in the set S_{i+1} and finally, increment i.
Step 6 Go to Step 3.

Now to walk through an example to illustrate the algorithm's steps. Let's take the knowledge base, $\{\bullet p \rightarrow \neg f, \bullet b \rightarrow f, p \rightarrow b\}$. We can read the statements as *"typical penguins do not fly"*, *"typical birds do fly"* and *"penguins are birds"*. Considering the statements we have, the situations we would desire the most are situations where there are no penguins while the most typical birds do fly as these would satisfy all the statements. It seems reasonable that the next best situation is when the most typical penguins do not fly while we can have that atypical birds also do not fly. Next, we have situations where atypical penguins do fly. The least desirable situations are when we have penguins that are not birds at all as this violates a classical conditional. Now we check if the model's reasoning matches our intuition. First, we note that because of the last statement we can immediately discount the valuations $\{p, \neg b, f\}$ and $\{p, \neg b, \neg f\}$ as having infinite rank as they will never satisfy the set of statements. We have the following valuations to examine: $\{\neg p, b, f\}, \{\neg p, \neg b, f\}, \{\neg p, \neg b, \neg f\}, \{p, b, \neg f\}, \{\neg p, b, \neg f\}, \{p, b, f\}$. So we begin by setting the rank of all the valuations to 0. The valuations that satisfy all the statements are $\{\neg p, b, f\}$, $\{\neg p, \neg b, f\}$ and $\{\neg p, \neg b, \neg f\}$. Therefore, they become the first level of our model, S_1, and this decreases the remaining valuations to check. $S_1 := [\![\mathcal{KB}]\!]^{\mathscr{R}_0} = \{\{\neg p, b, f\}, \{\neg p, \neg b, f\}, \{\neg p, \neg b, \neg f\}\}$. All the valuations not in S_1 obtain a rank of 1. The valuations that satisfy all the statements w.r.t. $\mathscr{R}_1$ are $S_2 := [\![\mathcal{KB}]\!]^{\mathscr{R}_1} = \{\{p, b, \neg f\}, \{\neg p, b, \neg f\}\}$. S_3 will be the valuation $\{p, b, f\}$. As previously mentioned, the valuations in S_4, which are $\{p, \neg b, f\}$ and $\{p, \neg b, \neg f\}$, will not satisfy the statements so S_4 will remain the same as S_5 and so on. The algorithm terminates at this stage. The algorithm's steps are both concisely and formally presented in Algorithm 4

$\mathscr{R}_0$	0.	$\{\neg p, b, f\}, \{\neg p, \neg b, f\}, \{\neg p, \neg b, \neg f\}, \{p, b, \neg f\}$ $\{\neg p, b, \neg f\}, \{p, b, f\}, \{p, \neg b, f\}, \{p, \neg b, \neg f\}$

$\mathscr{R}_1$	1.	$\{p, b, \neg f\}, \{\neg p, b, \neg f\}, \{p, b, f\}, \{p, \neg b, f\}, \{p, \neg b, \neg f\}$
	0.	$\{\neg p, b, f\}, \{\neg p, \neg b, f\}\{\neg p, \neg b, \neg f\}$

$\mathscr{R}_2$	2.	$\{p, b, f\}, \{p, \neg b, f\}, \{p, \neg b, \neg f\}$
	1.	$\{p, b, \neg f\}, \{\neg p, b, \neg f\}, \{p, b, f\}, \{p, \neg b, f\}, \{p, \neg b, \neg f\}$
	0.	$\{\neg p, b, f\}, \{\neg p, \neg b, f\}\{\neg p, \neg b, \neg f\}$

$\mathscr{R}_3$	3.	$\{p, \neg b, f\}, \{p, \neg b, \neg f\}$
	2.	$\{p, b, f\}$
	1.	$\{p, b, \neg f\}, \{\neg p, b, \neg f\}$
	0.	$\{\neg p, b, f\}, \{\neg p, \neg b, f\}\{\neg p, \neg b, \neg f\}$

$\mathscr{R}^*_{\mathcal{KB}}$	∞	$\{p, \neg b, f\}, \{p, \neg b, \neg f\}$
	2.	$\{p, b, f\}$
	1.	$\{p, b, \neg f\}, \{\neg p, b, \neg f\}$
	0.	$\{\neg p, b, f\}, \{\neg p, \neg b, f\}\{\neg p, \neg b, \neg f\}$

Figure 5.1: The ranked models for the Bird example generated during the execution of the LM-entailment algorithm. $\mathscr{R}^*_{\mathcal{KB}}$ is the final model and gives us the entailment.

Definition 5.3. ***(LM-entailment)*** *Let's say we have $\mathcal{KB} \subseteq \mathcal{L}^{\bullet}$ and $a \in \mathcal{L}^{\bullet}$, then $\mathcal{KB} \mathrel{|\!\approx}_{LM} a$, which means $\mathcal{KB}$* ***LM-entails*** *a, if $\mathscr{R}^*_{\mathcal{KB}} \Vdash a$. $\mathscr{R} \Vdash a$ denotes that $\mathscr{R}$ is a ranked model of a and is true if $[\![a]\!]^{\mathscr{R}} = \mathcal{W}^{\mathscr{R}}$, which tells* *us that the valuations that satisfy α are equal to all the possible valuations in $\mathscr{R}$.*

.

Algorithm 4: LM-entailment

```
Input: $\mathcal{KB}$
Output: $\mathscr{R}^*_{\mathcal{KB}}$
$\Phi_{\mathcal{KB}} := \{p \mid p$ is a propositional letter in $\mathcal{KB}\}$;
$W$ is the non-empty set of possible valuation for $\Phi_{\mathcal{KB}}$;
$\mathscr{R}_0(v) := 0$ for every $v \in W$;
$S_0 := \emptyset$;
$S_1 := [\![\mathcal{KB}]\!]^{\mathscr{R}_0}$;
i := 1;
while $S_i \neq S_{i-1} < 0$ do
    $\mathscr{R}_i := (\mathscr{R}_{i-1})^1_{S_i}$;
    $S_{I+1} := [\![\mathcal{KB}]\!]^{\mathscr{R}_i}$;
    $i := i + 1$;
$\mathscr{R}^*_{\mathcal{KB}} := (\mathscr{R}_i)^{\infty}_{S_i}$;
return $\mathscr{R}^*_{\mathcal{KB}}$
```

5.4 PT-entailment

This section details the second entailment method for PTL, PT-entailment, which is the shortened term for *Presumption of Typicality*-entailment [1], referring to the property previously detailed in section 3.6. The intuition behind this entailment corresponds with that of the *Presumption of Typicality* property. For this entailment, we consider the models in which each valuation is taken to be as low or as typical as possible with respect to the satisfaction of the knowledge base. These will be the models in $min_{\trianglelefteq_{PT}} Mod(\mathcal{KB})$ where $Mod(\mathcal{KB}) := \{\mathscr{R} \mid \mathscr{R} \Vdash \bigwedge \mathcal{KB}\}$. The relation $\trianglelefteq_{PT}$ is defined as follows.

Definition 5.4. *For two ranked interpretations $\mathscr{R}_1$ and $\mathscr{R}_2$, $\mathscr{R}_1 \trianglelefteq_{PT} \mathscr{R}_2$ if and only if for every $v \in W$, $\mathscr{R}_1(v) \leq \mathscr{R}_2(v)$, where $\mathscr{R}_1(v)$ represents the rank of the valuation v in $\mathscr{R}_1$. We also have that $\mathscr{R}_1 \triangleleft_{PT} \mathscr{R}_2$ if and only if $\mathscr{R}_1 \trianglelefteq_{PT} \mathscr{R}_2$ and not $\mathscr{R}_2 \trianglelefteq_{PT} \mathscr{R}_1$. Note that the relation $\trianglelefteq_{PT}$ is a preorder over the ranked interpretations.*

This entailment can produce several minimal models as opposed to the single model of LM-entailment and a formula is entailed by $\mathcal{KB}$ if and only if it is true in all the minimal models of $\mathcal{KB}$. Note that the single model produced by LM-entailment will always be among these minimal models [1]. This means that PT-entailment infers, at most, what is inferred by the LM-entailment model since we can base it on multiple models on top of the LM-entailment model. PT-entailment cannot come up with conclusions that are not entailed by LM-entailment. The following example is presented by Booth et al. [1]. The knowledge base is $\{\bullet\top \rightarrow (\neg p \wedge \neg r), \bullet p \rightarrow \bullet\neg f, \bullet r \rightarrow \bullet f, p \rightarrow \neg r\}$. We read the statements as *"the most typical things are neither penguins nor robins"*, *"typical penguins are typical non-flying birds"*, *"typical robins are typical flying birds"* and *" penguins are not robins"*. The ranked model, $\mathscr{R}^*_{\mathcal{KB}}$, without the ∞-rank valuations is:

0.	$\{\neg f, \neg p, \neg r\}, \{f, \neg p, r\}$

.

The minimality concept of PT-entailment then gives us the following three models where LM-entailment's $\mathscr{R}^*_{\mathcal{KB}}$ is the ranked interpretation $\mathscr{R}_1$ amongst these models.

$\mathscr{R}_1$	0.	$\{\neg f, \neg p, \neg r\}, \{f, \neg p, r\}$

$\mathscr{R}_2$	2.	$\{f, p, \neg r\}$
	1.	$\{\neg f, \neg p, \neg r\}, \{\neg f.p, \neg r\}$
	0.	$\{f, \neg p, \neg r\}$

$\mathscr{R}_3$	2.	$\{\neg f, \neg p, r\}$
	1.	$\{f, \neg p, r\}, \{f.\neg p, \neg r\}$
	0.	$\{\neg f, \neg p, \neg r\}$

Figure 5.2: The three PT-entailment models for the knowledge base $\{\bullet\top \rightarrow (\neg p \wedge \neg r), \bullet p \rightarrow \bullet\neg f, \bullet r \rightarrow \bullet f, p \rightarrow \neg r\}$

For a statement to be true, it must be satisfied by all three models. For example, we can derive that the statement $\bullet f \rightarrow \neg p$ is entailed as all the lowest valuations where f is satisfied also satisfy $\neg p$ in all three models. This derived statement tells us that the most typical flying things are not penguins, which is a sensible conclusion.

Definition 5.5. ***(PT-entailment)*** *Let's say we have* $\mathcal{KB} \subseteq \mathcal{L}^{\bullet}$ *and* $\alpha \in \mathcal{L}^{\bullet}$*, then* $\mathcal{KB} \approx_{PT} \alpha$*, which means* $\mathcal{KB}$ ***PT-entails*** α *if and only if* $min_{\trianglelefteq_{PT}} Mod(\mathcal{KB}) \subseteq Mod(\alpha)$*.*

Chapter 6

Deontic Analysis Process

This chapter briefly details the process used to check the effectiveness of the logic systems we are investigating. This efficacy refers to how well the logic system can be used to represent and reason about the deontic scenarios. Much of the evaluation of such efficacy will rely on intuition gained through assessing the literature. There are many discussions and results, which speak to what is deemed to be acceptable and reasonable, that can be leaned on as intuitive guidance. The analysis process for each logic system broadly involves working through the following three stages: Representation, Properties and Paradoxes. We discuss these stages now in more detail.

Representation

First, we must settle upon a translation of deontic statements into the specific language of the logic being analysed. We must be able to consistently apply this representation method and it must provide an intuitive reading. This reading is guided by the semantics of the respective logics, as statements in their languages have established interpretations. Once we decide upon this method of representation, the process would continue to the next stage which deals with the properties.

Properties

This stage investigates which of the deontic properties are applicable in the logic that is being analysed. Ideally, the logic in question would generally satisfy each property. If that is not the case, we check whether the property can be reasonably applied to translated statements in some instances. This process will comprise assuming the property's premises and observing consequences that followed. We have the following to check: Restricted Strengthening of the Antecedent, Factual Detachment, Deontic Detachment, Transitivity, Disjunction, Conjunction, Weakening, Ought Implies Can principle and Distribution along with the satisfaction of the Deontic Explosion Principle. Once we check the properties within the logic system, we continue to the analysis of the paradoxes since we now have a way to represent their statements and the properties we can use to perform derivations.

Paradoxes

To analyse the paradoxes, we first translate the statements of each paradox into the relevant logic's language and then we take the paradox through the available reasoning methods. When interpreting the paradoxes and their results, there are two options in terms of interpretations that can be made when confronted with conflicting obligations [27].

1. We could accept the derivation of $\bigcirc\bot$ which explicitly tells us that there is a dilemma with the obligations. The reasoning methods would need to accommodate this, allowing us to explicitly state that there is a violation while continuing to produce other sensible derivations.

2. We could accept the "ought implies can" principle which tells us that the derivation of $\bigcirc\bot$ is undesirable since we do not want to be obligated to perform two contradictory tasks. In this situation, we would want to derive the best possible obligations in the less-than-ideal scenario where we have a conflict.

Much of the literature suggests that the second approach is the most reasonable interpretation and since it is the most sensible to us, it will be the one we adopt moving forward [21, 23]. We restate the paradoxes that will be analysed: Forrester's paradox, Chisholm's paradox, Fence paradox, the Trump/Kim paradox and Van Fraassen's paradox. With PTL-entailment, where we cannot analyse the paradoxes for reasons that will outlined in chapter 9, we put forward alternative examples. These alternative examples are meant to replicate much of the intuition displayed by the paradoxes.

Chapter 7

Deontic Analysis Using KLM-style Defeasible Reasoning

This chapter presents the analysis of KLM-style defeasible reasoning and its effectiveness in handling deontic scenarios. KLM-style defeasible reasoning was formally presented in chapter 3 and the analysis process was outlined in the previous chapter. We begin by establishing a defeasible translation for the deontic statements using the defeasible implication operator $\mid\!\sim$. Then the properties section shows which of the desired deontic properties are satisfied by the KLM approach. The motivation to perform the analysis of the paradoxes using the lexicographic closure algorithm, rather than that of rational closure, is then briefly presented. Thereafter, we observe how we can use lexicographic closure to block the undesirable derivations of the paradoxes.

7.1 Representation

In order to select a method with which to represent deontic statements using the KLM defeasible implications, we take a look at the similar semantics of DSDL and the KLM approach. Recall that a deontic obligation such as $\bigcirc(\beta \mid \alpha)$ tells us that the 'best' α-valuations are also β-valuations, and $\alpha \mid\!\sim \beta$ has the same reading. This is to say that given the set of valuations that satisfy α, $[\![\alpha]\!] = \{s \in V : s \models \alpha\}$, both $\bigcirc(\beta \mid \alpha)$ and $\alpha \mid\!\sim \beta$ provide us with similar interpretations. The former states that $\forall\, s' \in V$, if $s' \in \{s \in [\![\alpha]\!] : s \leq t, \forall\, t \in [\![\alpha]\!]\}$ where $\leq$ is a total preorder over W, then $s' \models \beta$. The latter, on the other hand, states that $\forall\, s' \in V$, if $s' \in \{s \in [\![\alpha]\!] : \nexists t \in$

$[\![\alpha]\!]$, *such that* $t < s\}$, then $s' \models \beta$, where $\leq$ is a modular order. We present a guide to translate the deontic statements into their KLM equivalents. For the translation of deontic obligations into the language of the KLM approach, we say that $\bigcirc(\beta \mid \alpha)$ is equivalent to $\alpha \mid\sim \beta$. Non-conditional obligations such as $\bigcirc\alpha$ will be represented as $\top \mid\sim \alpha$. Facts are given in the usual way, such as α. This gives an appropriate method to handle the translation of deontic statements and allows for the analysis to continue.

7.2 Properties

In the following section, we look at which of the deontic properties that we deemed desirable are satisfied by the KLM-style defeasible reasoning approach. This begins with an analysis of the "ought implies can" principle's satisfaction and its subsequent impact on how we reason within the KLM logic system.

7.2.1 Ought Implies Can Principle and Violations

When working with obligations, we ideally want to be able to explicitly state whether a certain fact brings up a violation with respect to our set of obligations. This will require the logic system we use to cater for the occurrence of conflicts. But when dealing with these KLM defeasible implications we will see that neither of our chosen KLM reasoning algorithms, rational closure nor lexicographic closure, can perform their reasoning with the presence of conflicts and thus satisfy the "ought implies can" principle. If one is obligated to perform a task then they should be able to do so, therefore the task should not be simultaneously obligatory and prohibited. This principle can be represented by the derivation of $\neg \bigcirc (\alpha \wedge \neg\alpha)$ which tells us that there is no obligation to perform contradictory tasks. With conflicts being an explicit indication that a violation has occurred, the "ought implies can" principle means we must find a way to detect violations.

Recall that in many of the paradoxes, there are facts which represent the occurrence of a violation of some obligation. For example, amongst Forrester's paradox's statements, we had the non-conditional obligation $\bigcirc\neg k$ and the fact k. The latter representing a violation of the former. As a result of the "ought implies can" principle, when we analyse contrary-to-duty obligations we cannot have the contradictory facts in the defeasible knowledge base as this will cause conflicts. This is due to the need for the defeasible knowledge base to be consistent when using the reasoning algorithms. Instead, we will initially remove the facts from the knowledge base and determine which general obligations arise from the set of obligations we have along with the background knowledge. We then use the facts as a premise when examining the derivations of actual obligations. These actual obligations refer to the obligations which a hypothetical agent must act upon when we take certain facts into consideration. For example, let's say we have the following set of statements $\{p, \bigcirc\neg p, \bigcirc q, \bigcirc(r \mid p)\}$ which are in deontic form. We will remove the fact p from the knowledge base and observe that we initially have the obligations $\bigcirc\neg p$, $\bigcirc q$ and $\bigcirc(r \mid p)$. Once we reintroduce p, we see that we no longer have the obligation $\bigcirc\neg p$ but now derive the non-conditional obligation $\bigcirc r$, derived using the Deontic Detachment property, amongst the other remaining obligations.

This approach to dealing with these contradictory instances allows us to generate conclusions from the facts and knowledge base. The intuition of this approach is that when we do not have the fact as part of the knowledge base, derivations can be thought of as occurring in a world where this fact has not transpired. We consider these to be general or primary obligations. Once we introduce a fact, derivations will be interpreted as if the fact holds and thus obligations that were formally in force may no longer be, given this fact. And since we cannot have conflicts this immediately blocks deontic explosion which is the desired consequence.

A consequence of avoiding conflicts in order to do reasoning is that we cannot explicitly detect specific violations during the use of the reasoning algorithms. The reasoning methods we use can give us the best-case scenario in an undesirable situation that a violation has occurred but there is no way to state whether an obligation was violated during reasoning. To circumvent this, we temporarily keep a fact in the knowledge base before the application of the reasoning algorithm and check if the knowledge base is inconsistent. If so, then we must determine which obligations have been violated. To do this, we test for every obligation in the knowledge base, whether this obligation is in conflict with the fact in question. All those obligations which form an inconsistent pair with the fact will be those which have been violated.

We note another drawback to this approach on facts. This being the inability to represent statements that specifically tell us whether an obligation has been fulfilled or an obligation has been violated. An example of this being $(\bigcirc p) \wedge p$ to explicitly state that the obligation to do p has been fulfilled or the violation-equivalent $(\bigcirc p) \wedge \neg p$ [20].

7.2.2 Remaining properties

In chapter 3 we have that the following desired deontic properties are satisfied by the defeasible entailment relation of both rational closure and lexicographic closure, $\mathrel{|\!\!\approx}$: Weakening, Conjunction, Disjunction and a version of Transitivity called Cumulative Transitivity. We restate these properties below and show that the following properties are also satisfied: Distribution, Factual Detachment and Deontic Detachment.

Weakening

$$\frac{\models \beta \rightarrow \gamma, \mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \beta}{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \gamma}$$

Conjunction

$$\frac{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \beta, \mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \gamma}{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \beta \wedge \gamma}$$

Disjunction

$$\frac{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \gamma, \mathcal{KB} \mathrel{|\!\!\approx} \beta \mathrel{|\!\!\sim} \gamma}{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \vee \beta \mathrel{|\!\!\sim} \gamma}$$

Cumulative Transitivity

$$\frac{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \beta, \mathcal{KB} \mathrel{|\!\!\approx} \alpha \wedge \beta \mathrel{|\!\!\sim} \gamma}{\mathcal{KB} \mathrel{|\!\!\approx} \alpha \mathrel{|\!\!\sim} \gamma}$$

Distribution

The property of Distribution was presented in section 4.3 as follows.

$$\text{If} \models \alpha \rightarrow \beta, \text{ then } \mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\alpha \rightarrow \bigcirc\beta$$

Recall that this tells us that if we assume that α implies β, then if we have that one is obligated to do α then one is obligated to do β. We represent it using the following KLM derivation.

$$\frac{\models \alpha \rightarrow \beta, \mathcal{KB} \mathrel{|\!\!\approx} \top \mathrel{|\!\!\sim} \alpha}{\mathcal{KB} \mathrel{|\!\!\approx} \top \mathrel{|\!\!\sim} \beta}$$

This property is satisfied, and we can see this if we apply the Weakening property. Assume we have $\alpha \rightarrow \beta$ as a tautology and the statement $\top \mathrel{|\!\!\sim} \alpha$ which represents $\bigcirc\alpha$. The application of the Weakening property then enables the derivation of $\top \mathrel{|\!\!\sim} \beta$, which represents $\bigcirc\beta$.

Factual Detachment

The following is the property of Factual Detachment in deontic form and we want to show that it holds for the KLM approach.

$$\frac{\mathcal{KB} \approx\!\!\!\mid \bigcirc(\beta \mid \alpha), \models \alpha}{\mathcal{KB} \approx\!\!\!\mid \bigcirc\beta}$$

The following is the property translated into the language of the KLM approach.

$$\frac{\mathcal{KB} \approx\!\!\!\mid \alpha \mathrel{|\!\sim} \beta, \models \alpha}{\mathcal{KB} \approx\!\!\!\mid \top \mathrel{|\!\sim} \beta}$$

If we have that α is a tautology then we can state $\alpha \mathrel{|\!\sim} \beta$ as $\top \mathrel{|\!\sim} \beta$ so we have that the property holds.

Deontic Detachment

The following is the property of Deontic Detachment in deontic form and we want to show that it holds for the KLM approach.

$$\frac{\mathcal{KB} \approx\!\!\!\mid \bigcirc(\beta \mid \alpha), \mathcal{KB} \approx\!\!\!\mid \bigcirc\alpha}{\mathcal{KB} \approx\!\!\!\mid \bigcirc\beta}$$

The following is the property translated into the language of the KLM approach.

$$\frac{\mathcal{KB} \approx\!\!\!\mid \alpha \mathrel{|\!\sim} \beta, \mathcal{KB} \approx\!\!\!\mid \top \mathrel{|\!\sim} \alpha}{\mathcal{KB} \approx\!\!\!\mid \top \mathrel{|\!\sim} \beta}$$

Let's say we have $\{\alpha \mathrel{|\!\sim} \beta, \top \mathrel{|\!\sim} \alpha\}$ and we want to derive $\top \mathrel{|\!\sim} \beta$. It holds that $\alpha \equiv \top \wedge \alpha$, so by Left Logical Equivalence we obtain $\top \wedge \alpha \mathrel{|\!\sim} \beta$. If we take that derived defeasible implication, $\top \wedge \alpha \mathrel{|\!\sim} \beta$, and combine it with $\top \mathrel{|\!\sim} \alpha$ then we can use the Cumulative Transitivity property to get $\top \mathrel{|\!\sim} \beta$.

7.2.3 Strengthening of the Antecedent

Recall the Strengthening of the Antecedent property from deontic logic presented in chapter 4, which is given in deontic form below.

$$\frac{\mathcal{KB} \approx\!\!\!\mid \bigcirc(\beta \mid \alpha)}{\mathcal{KB} \approx\!\!\!\mid \bigcirc(\beta \mid \alpha \wedge \gamma)}$$

Since the reasoning we will use is non-monotonic, we will not have the above property of Strengthening of the Antecedent/Monotonicity [13]. We instead have the property of Rational Monotonicity with KLM-style defeasible reasoning.

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

We now attempt to convert this property into deontic form in order to determine whether it reads suitably and can, therefore, be adopted as an alternative to Strengthening of the Antecedent for the KLM approach. The KLM statement $\alpha \mid\!\sim \beta$ can be translated to $\bigcirc(\beta \mid \alpha)$ and the derived KLM statement $\alpha \wedge \gamma \mid\!\sim \beta$ can be translated to $\bigcirc(\beta \mid \alpha \wedge \gamma)$. Now to translate the KLM statement $\alpha \not\mid\!\sim \neg\gamma$. This statement tells us that it is not the case that every minimal valuation that satisfies α also satisfies $\neg\gamma$. In other words, there exists at least one minimal α-valuation that satisfies γ. To translate it, we would need a deontic statement that states that there exists a minimal valuation that satisfies α that does not satisfy $\neg\gamma$. This can be written in deontic form as $\mathcal{KB} \approx\!\!\mid\!\!\;\, \neg \bigcirc (\neg\gamma \mid \alpha)$. Thus the deontic version of the entire translated property appears as follows.

$$\frac{\mathcal{KB} \approx\!\!\mid\!\!\;\, \bigcirc(\beta \mid \alpha), \mathcal{KB} \approx\!\!\mid\!\!\;\, \neg \bigcirc (\neg\gamma \mid \alpha)}{\mathcal{KB} \approx\!\!\mid\!\!\;\, \bigcirc(\beta \mid \alpha \wedge \gamma)}$$

Although not having the Strengthening of the Antecedent property blocks us from intuitive derivations, its omission also results in many issues within the paradoxes no longer arising, which we will detail in section 7.4. As seen in section 4.4, where the paradoxes were first presented, some of the issues within the paradoxes are a result of derivations made using a version of this Strengthening of the Antecedent property. Once we remove the property, those derivations are no longer possible.

7.3 Why use lexicographic closure?

In terms of the reasoning methods, we select from the two options previously presented in chapter 3. These two options are rational closure and lexicographic closure. The following is an example meant to justify the selection of lexicographic closure as the method of use from now on. The statements for our example are *"Students usually do not pay taxes"*, *"Students who work usually have to pay taxes"*, and *"Student usually have to be registered"*. We also have the statement *"Students who work are students"*, which acts as a fact so is therefore omitted from the following set.

$$\{s \mid\sim \neg t, s \wedge w \mid\sim t, s \mid\sim r\}$$

This is similar to the Bird example from chapter 3. This example's fact, *"Students who work are students"*, is analogous to the Bird example's $p \rightarrow b$ which is *"Penguins are birds"*, would be represented as $s \wedge w$. The statement $b \mid\sim f$ is analogous to $s \mid\sim \neg t$ while $p \mid\sim \neg f$ is analogous to $s \wedge w \mid\sim t$ and $b \mid\sim w$ is analogous to $s \mid\sim r$. It seems reasonable to infer that students who work must be registered, $s \wedge w \mid\sim r$, but this is not possible with rational closure. Whereas with lexicographic closure it is possible to make that derivation, even though students who work are not considered typical students. This is similar to how we can derive that penguins have wings in the Bird example using lexicographic closure, whereas rational closure fails to do so. The desire is for obligations which are usually in force for students to be applicable to atypical students unless we are given evidence to the contrary. These are the sort of derivations which justify the use of the less conservative, presumptive reading of lexicographic closure when doing normative reasoning. Our claim is not that the presumptive reading will always be more suitable than the prototypical reading, but it is the preferred choice moving forward with our deontic analysis.

7.4 Paradoxes

We proceed to apply the lexicographic closure algorithm to the paradoxes which will be translated using our chosen KLM representation.

7.4.1 Forrester's paradox

This paradox comprises three statements. The obligations *"You must not kill anybody"* and *"If you kill someone then you must kill them gently"* as well as the fact *"You killed someone"*. Along with these we also have the background knowledge, *"Killing gently implies killing"*. The KLM equivalent of these obligations is given in the following set of KLM statements and this comes with the translated background knowledge, $g \rightarrow k$ and the fact k.

$$\{\top \mid\sim \neg k, k \mid\sim g\}$$

The fact k causes the knowledge base to become inconsistent if we include it during reasoning. This means we have a violation. If we check the fact against every obligation in the materialisation set $\overrightarrow{\mathcal{KB}} = \{\top \rightarrow \neg k, k \rightarrow g\}$, it is clear that $\top \mathrel{|\!\sim} \neg k$ is the obligation that is violated. We proceed to determine the ranks of the statements. We now have a consistent $\overrightarrow{\mathcal{KB}}$ to construct the ranking of the statements. We have $\tilde{\mathcal{B}} = \{\top \mathrel{|\!\sim} \neg k, k \mathrel{|\!\sim} g\}$ with $r(\top \mathrel{|\!\sim} \neg k) = 0$ and $r(k \mathrel{|\!\sim} g) = 1$ so the ranking would be the following:

1	$k \mathrel{|\!\sim} g$
0	$\top \mathrel{|\!\sim} \neg k$

Now we examine what we can gather given the above ranking and compare this with the derivations shown in chapter 4. In that chapter, we observed the problematic conclusions that we could derived with the paradox's obligations and certain deontic properties.

Restricted Strengthening of the Antecedent, Weakening and Conjunction

1.	$\bigcirc \neg k \rightarrow_W \bigcirc \neg g$
2.	$\bigcirc \neg g \rightarrow_{RSA} \bigcirc(\neg g \mid k)$
3.	$\{\bigcirc(\neg g \mid k), \bigcirc(g \mid k)\} \rightarrow_{Conj} \bigcirc(\neg g \wedge g \mid k)$

We first want to check if we can derive $\bigcirc\neg g$ as a general obligation, via the Weakening property, before we apply the fact. As a non-conditional obligation, we assume the tautology, $\top$ as the premise and observe if we can derive $\neg g$. Recall that in chapter 3, the structure of the lexicographic closure derivation for a formula, let's say w, are $p \cup \mathcal{A} \cup \mathcal{D} \models w$ for every $\mathcal{D} \in \mathcal{T}$ where p is the premise. As stated above, we have $\top$ as the premise. We then have $\mathcal{D} = \{\top \rightarrow \neg k, k \rightarrow g\}$ and we include the background knowledge $g \rightarrow k$ in the set of statements which have infinite rank, $\mathcal{A}$, where it is the only statement. We have $\mathcal{A} = \{g \rightarrow k\}$ and note that the statement, $g \rightarrow k$, is equivalent to $\neg k \rightarrow \neg g$ from contraposition. We have the following derivation which shows that $\neg g$ follows from our knowledge base using lexicographic closure.

$$\{\top\} \cup \{g \rightarrow k\} \cup \{\top \rightarrow \neg k, k \rightarrow g\} \models \neg g$$

We move on to check whether we can derive the undesirable conditional obligation $\bigcirc(\neg g \mid k)$. We derived this using Restricted Strengthening of the Antecedent in the above derivations table, but we do not have this property. It is thus interesting to observe whether the alternative, Rational Monotonicity, ‘blocks’ this undesirable obligation from being derived. We would consider to be an actual obligation which occurs when we take the fact that killing has occurred. For this situation, we have k as the premise and $\mathcal{D} = \{k \rightarrow g\}$. This results in the following derivation which tells us we cannot derive $\neg g$ from the knowledge base given the fact k.

$$\{k\} \cup \{g \rightarrow k\} \cup \{k \rightarrow g\} \not\models \neg g$$

Intuitively, this tells us that although we can originally derive $\bigcirc \neg g$ which tells us *“do not kill gently”* in general, once one kills then we retract that obligation. The individual will now be obligated to kill gently once they have killed and this does which seems reasonable initially. But violated obligations no longer being derivable, referred to as the drowning problem, is considered an undesirable result in the literature [23]. It is important to note this drawback.

Factual Detachment and Conjunction

1.	$\{\bigcirc(g \mid k), k\} \rightarrow_{FD} \bigcirc g$
2.	$\{\bigcirc \neg k, \bigcirc g\} \rightarrow_{Conj} \bigcirc(\neg k \wedge g)$

For the first derivation, we examine whether we can derive $\bigcirc g$ given the fact k. With this being a non-conditional obligation, we would have normally have the tautology as the premise but taking the fact k into account means we use k as a premise. We also have that $\mathcal{D} = \{k \rightarrow g\}$ and $\mathcal{A} = \{g \rightarrow k\}$. The following derivation tells us we can derive $\bigcirc g$.

$$\{k\} \cup \{g \rightarrow k\} \cup \{k \rightarrow g\} \models g$$

We also see lexicographic closure blocks the issue that arises between Factual Detachment and Conjunction where $\bigcirc(\neg k \wedge g)$ can be derived. We continue to take k as the premise since it has already been taken it into account. We still have $\mathcal{D} = \{k \rightarrow g\}$ and $\mathcal{A} = \{g \rightarrow k\}$. Although g follows from this lexicographic closure configuration, we cannot derive $\neg k$ and therefore cannot derive the undesirable $\bigcirc(\neg k \wedge g)$ using lexicographic closure.

$$\{k\} \cup \{g \rightarrow k\} \cup \{k \rightarrow g\} \not\models \neg k \wedge g$$

This tells us that although we originally had the non-conditional obligation to not kill, if one kills then we can retract that non-conditional obligation, avoiding a violation. Then we can aim for the next best scenario, which would be for one to kill gently.

7.4.2 Chisholm's paradox

This paradox comprises four statements. The obligations *"Jones must go assist his neighbour"*, *"If Jones goes to assist his neighbour then he must tell them that he is coming"* and *"If Jones does not go to assist his neighbour then he must not tell them that he is coming"* as well as the fact *"Jones does not go to assist his neighbour"*. The following is the KLM-style representation of the paradox's statements. We have the following knowledge base along with the fact $\neg a$.

$$\{\top \mid\!\sim a, a \mid\!\sim t, \neg a \mid\!\sim \neg t\}$$

Introducing the fact into the knowledge base would lead to a conflict while using lexicographic closure and this is because of the obligation, $\top \mid\!\sim a$. This tells us that the obligation, *"Jones must go assist his neighbour"*, was violated by the fact of Jones not going to assist his neighbour. We construct the rankings with the knowledge base without the fact, $\neg a$. The set of defeasible implications with a rank less then infinity is $\tilde{\mathcal{B}} = \{\top \mid\!\sim a, a \mid\!\sim t, \neg a \mid\!\sim \neg t\}$ and the ranking is given below. We have $r(a \mid\!\sim t) = 0, r(\top \mid\!\sim a) = 0$ and $r(\neg a \mid\!\sim \neg t) = 1$.

1	$\neg a \mid\!\sim \neg t$
0	$\top \mid\!\sim a, a \mid\!\sim t$

Now we examine what can be gathered given the above ranking and compare this with the derivations shown in chapter 4.

Transitivity and Restricted Strengthening of the Antecedent

1.	$\{\bigcirc(t \mid a), \bigcirc a\} \rightarrow_{Trans} \bigcirc t$
2.	$\bigcirc t \rightarrow_{RSA} \bigcirc(t \mid \neg a)$
3.	$\{\bigcirc(t \mid \neg a), \bigcirc(\neg t \mid \neg a)\} \rightarrow_{Conj} \bigcirc(t \wedge \neg t \mid \neg a)$

The general obligation $\bigcirc t$, via Transitivity in the above table and can also be done via the Deontic Detachment property, is now shown to be derivable using lexicographic closure. Since we have not taken the fact into account, we use $\top$ as the premise, we have the following derivation of t with $\mathcal{D} = \{\top \rightarrow a, a \rightarrow t, \neg a \rightarrow \neg t\}$.

$$\{\top\} \cup \{\top \rightarrow a, a \rightarrow t, \neg a \rightarrow \neg t\} \models t$$

With $\bigcirc t$ being determined as a derivable general obligation, we then observe whether we can derive the conditional obligation $\bigcirc(t \mid \neg a)$. This was derived via the use of the Restricted Strengthening of the Antecedant property in the above table. We take the fact $\neg a$ to be true and then have that there is a single most serious subset which satisfies the premise, $\mathcal{D} = \{a \rightarrow t, \neg a \rightarrow \neg t\}$. But if we take $\neg a$ to be the premise, we then have the following lexicographic closure derivation.

$$\{\neg a\} \cup \{a \rightarrow t, \neg a \rightarrow \neg t\} \not\models t$$

This tells us that the derivation of $\bigcirc(t \mid \neg a)$ is blocked. We derived this from $\bigcirc t$ and the fact $\neg a$ using Restricted Strengthening of the Antecedent. With lexicographic closure not having the Restricted Strengthening of the Antecedent property, it seems sensible that the undesirable derivations from this property are then blocked.

Deontic Detachment, Factual Detachment and Conjunction

1.	$\{\bigcirc(t \mid a),\ \bigcirc a\} \rightarrow_{DD} \bigcirc t$
2.	$\{\bigcirc(\neg t \mid \neg a),\ \neg a\} \rightarrow_{FD} \bigcirc \neg t$
3.	$\{\bigcirc t, \bigcirc \neg t\} \rightarrow_{Conj} \bigcirc t \wedge \neg t$

We now see that the issue that arises with the Factual Detachment and Deontic Detachment properties being used together, is no longer present. The first derivation, of $\bigcirc t$ using Deontic Detachment, is satisfied and is identical to the derivation via Transitivity seen in section 7.4.1. $\bigcirc t$ is an obligation that we consider before taking the fact $\neg a$ into account, so we use $\top$ as the premise. We have $\mathcal{D} = \{\top \rightarrow a, a \rightarrow t, \neg a \rightarrow \neg t\}$ and this gives the following lexicographic closure derivation.

$$\{\top\} \cup \{\top \rightarrow a, a \rightarrow t, \neg a \rightarrow \neg t\} \models t$$

This result shows us that $\bigcirc t$ can be derived in general. We now examine whether $\bigcirc \neg t$ follows when we take the fact $\neg a$ as having occurred. And with $\neg a \rightarrow \neg t$ being in the single most serious subset, $\mathcal{D} = \{a \rightarrow t, \neg a \rightarrow \neg t\}$, this allows the derivation of $\neg t$.

$$\{\neg a\} \cup \{a \rightarrow t, \neg a \rightarrow \neg t\} \models \neg t$$

We now check whether we can derive t if we continue to use $\neg a$ as a premise. It remains that $\mathcal{D} = \{a \rightarrow t, \neg a \rightarrow \neg t\}$ and t does not follow as seen in the below derivation. This prevents $t \wedge \neg t$ from being derived when we have the fact $\neg a$.

$$\{\neg a\} \cup \{a \rightarrow t, \neg a \rightarrow \neg t\} \not\models t \wedge \neg t$$

The intuition we now have is that the non-conditional obligation $\bigcirc t$, which read as *"Jones should tell his neighbour that he is coming"*, is no longer in force when the fact $\neg a$ has happened.

7.4.3 Fence paradox and exceptions

This paradox comprises four statements. The obligations *"There must not be a fence"*, *"If there is a fence then it must be a white fence"* and *"If there is a dog then there must be a white fence"* as well as the fact *"There is a fence"*. This paradox illustrates the need to distinguish between exceptions and violations in contrary-to-duty scenarios. We have the following knowledge base with the fact f.

$$\{\top \mid\!\sim \neg f, f \mid\!\sim (f \wedge w), d \mid\!\sim (f \wedge w)\}$$

The general obligation $\top \mid\!\sim \neg f$, *"There must not be a fence"*, is violated when we assume the fact f. We want it to be the case that when a white fence is present, meaning that $f \wedge w$ holds, that we get that there should be a dog, d should hold. But we cannot get this derivation with the current KLM representation of the deontic statements as we can see below. We have the set $\tilde{\mathcal{B}} = \{\top \mid\!\sim \neg f, f \mid\!\sim (f \wedge w), d \mid\!\sim (f \wedge w)\}$ and we give the ranking of these statements.

1	$f \mid\!\sim (f \wedge w), d \mid\!\sim (f \wedge w)$
0	$\top \mid\!\sim \neg f$

If we take $f \wedge w$ to be the premise, the single most serious subset that satisfies $f \wedge w$ is $\{f \rightarrow (f \wedge w), d \rightarrow (f \wedge w)\}$ which shows that the derivation of d is not possible using lexicographic closure.

$$\{f \wedge w\} \cup \{f \rightarrow (f \wedge w), d \rightarrow (f \wedge w)\} \not\models d$$

A representation of exceptions is not immediately obvious using KLM-style defeasible reasoning and must still be explored.

7.4.4 Trump/Kim paradox

This paradox comprises four obligations. *"Trump must not be told the secret"*, *"Kim must not be told the secret"*, *"If you tell Trump then you must tell Kim"* and *"If you tell Kim then you must tell Trump"*. The KLM version of the set of obligations is now given.

$$\{\top \mid\!\sim \neg t, \top \mid\!\sim \neg k, t \mid\!\sim k, k \mid\!\sim t\}$$

The entire knowledge base makes up the set of non-infinitely ranked obligations, $\tilde{\mathcal{B}}$, and we have the following ranking.

1	$t \mid\!\sim k, k \mid\!\sim t$
0	$\top \mid\!\sim \neg t, \top \mid\!\sim \neg k$

Restricted Strengthening of the Antecedent, Weakening and Conjunction

1.	$\{\bigcirc\neg t, \bigcirc\neg k\} \rightarrow_{Conj} \bigcirc(\neg t \wedge \neg k)$
2.	$\bigcirc(\neg t \wedge \neg k) \rightarrow_{W} \bigcirc\neg(t \wedge k)$
3.	$\bigcirc\neg(t \wedge k) \rightarrow_{RSA} \bigcirc(\neg(t \wedge k) \mid t)$
4.	$\{\bigcirc(\neg(t \wedge k) \mid t), \bigcirc(t \wedge k \mid t\} \rightarrow_{Conj} \bigcirc(\neg(t \wedge k) \wedge t \wedge k \mid t)$

We have the following two derivations which represent, firstly, Conjunction used on the two non-conditional obligations and subsequently, Weakening applied to that resultant conjunction.

$$\{\top\} \cup \{\top \rightarrow \neg t, \top \rightarrow \neg k, k \rightarrow t, t \rightarrow k\} \models \neg t \wedge \neg k$$

$$\{\top\} \cup \{\top \rightarrow \neg t, \top \rightarrow \neg k, k \rightarrow t, t \rightarrow k\} \models \neg(t \wedge k)$$

We now examine if we can derive the conditional obligation $\bigcirc(\neg(t \wedge k) \mid t)$. Once we use the fact t as the premise, the non-conditional obligations $\top \mid\!\sim \neg t$ and $\top \mid\!\sim \neg k$ are no longer in the most serious subset. $\top \mid\!\sim \neg k$ cannot be since we have $t \rightarrow k$ in the subset and this in combination with t will cause a conflict. And thus we cannot derive $\neg(t \wedge k)$ since the conditional obligation $t \rightarrow k$ gives as that k is true along with t. This is shown in the following derivation.

$$\{t\} \cup \{k \rightarrow t, t \rightarrow k\} \not\models \neg(t \wedge k)$$

This blocks the derivation of the conditional obligation $\bigcirc(\neg(t \wedge k) \wedge t \wedge k \mid t)$. Intuitively, this tells us that once we tell one of them the secret, the obligation to tell neither of them falls away and we must reveal the secret to the other.

7.4.5 Van Fraassen's paradox

This paradox comprises two obligations. *"You must honour your father or honour your mother"* and *"You must not honour your mother"*. The obligations can be represented by the following set of KLM statements.

$$\{\top \mid\!\sim f \vee m, \top \mid\!\sim \neg m\}$$

Weakening and Conjunction

1.	$\bigcirc\alpha \rightarrow_W \bigcirc(\alpha \vee \beta)$
2.	$\{\bigcirc(\alpha \vee \beta), \bigcirc\neg\alpha\} \rightarrow_{Conj} \bigcirc(\neg\alpha \wedge \beta)$
3.	$\bigcirc(\neg\alpha \wedge \beta) \rightarrow_W \bigcirc\beta$

The above undesirable derivation of deontic explosion is blocked because we have the "ought implies can" principle and thus cannot have both $\bigcirc\alpha$ and $\bigcirc\neg\alpha$ in the knowledge base. If we have $\tilde{\mathcal{B}} = \{\top \mid\!\sim f \vee m, \top \mid\!\sim \neg m\}$ then we can see that the two statements will have the same rank.

0	$\top \mid\!\sim f \vee m, \top \mid\!\sim \neg m$

We want to check if we can derive $\top \mid\!\sim f$ which would be *"You must honor your father"*. This is a non-conditional obligation so the premise will be the tautology, $\top$. And since both statements satisfy the tautology, we can clearly see that the intuitive *"You must honor your father"* can be derived.

$$\{\top\} \cup \{\top \rightarrow f \vee m, \top \rightarrow \neg m\} \models f$$

We now look at the alphabetic variant, the Apples-and-Pears example. This example has the obligations, *"You must buy apples or buy pears"* and *"You must not buy apples"*.

$$\{\top \mid\!\sim a \vee p, \top \mid\!\sim \neg a\}$$

In the Apples-and-Pears example, we could derive $\bigcirc(p \mid a)$ from $\bigcirc p$ through the use of Restricted Strengthening of the Antecedent. This is a derivation we wish to avoid since this does not fully align with our intuition.

Weakening, Conjunction and Restricted Strengthening of the Antecedent

1.	$\{\bigcirc a \vee p, \bigcirc \neg a\} \rightarrow_{Conj} \bigcirc \neg a \wedge p$
2.	$\bigcirc \neg a \wedge p \rightarrow_{W} \bigcirc p$
3.	$\bigcirc p \rightarrow_{RSA} \bigcirc (p \mid a)$

Since the Apples-and-Pears example is analogous to Van Fraassen's paradox, we know that $\bigcirc p$ is derivable from our knowledge base in the same way $\bigcirc f$ follows in Van Fraassen's paradox.

$$\{\top\} \cup \{\top \rightarrow p \vee a, \top \rightarrow \neg a\} \models p$$

We now check if we can block the derivation of $\bigcirc(p \mid a)$, $a \mid\!\sim p$ in the KLM representation, from the knowledge base. Take a to be the premise, we can see that $\top \rightarrow \neg a$ does not satisfy the premise and thus the most serious subset contains the lone obligation, $\top \rightarrow p \vee a$. We then see that p cannot be derived.

$$\{a\} \cup \{\top \rightarrow p \vee a\} \not\models p$$

This aligns with the contextual interpretation from chapter 4.4.5 which tells us that once apples have been bought it should be the case that pears ought to be bought as well.

7.5 Conclusion

We began by determining a method to use the KLM-style defeasible implications to represent deontic statements. The KLM equivalent of an obligation such as $\bigcirc(\beta \mid \alpha)$ would be $\alpha \mid\!\sim \beta$ while a non-conditional obligation $\bigcirc\alpha$ would be represented with $\top \mid\!\sim \alpha$ and facts would be represented in the usual way, such as α. KLM-style defeasible reasoning satisfies most of the deontic properties which are seen as suitable for a deontic system to have. The Restricted Strengthening of the Antecedent property was not satisfied because of the non-monotonicity of KLM, but there is the Rational Monotonicity property which we can use as an alternative. The results when looking at the paradoxes are satisfactory in that undesirable conclusions are not derivable using the lexicographic closure algorithm. There is also a way to identify which obligations are violated by a fact in the paradoxes, via the lexicographic closure algorithm's consistency check, but there is still a lack of an obvious representation for exceptions as seen in the Fence paradox.

Chapter 8

Deontic Analysis Using LM-entailment

In this chapter, we analyse the efficacy of PTL's LM-entailment reasoning method in the deontic domain. We start by detailing the chosen representation method and then investigate which of the deontic properties can be applied once converted to their PTL translations. The chapter then moves onto the analysis of the paradoxes which involves translating the paradoxes into PTL and applying the LM-entailment algorithm to produce the LM-entailment model which we can reason with.

8.1 Representation

Due to the expressiveness of PTL, settling on a representation for PTL statements was a harder task than finding the representation with the KLM language. The following sections present the representation and the intuition behind its selection. If there is a deontic scenario that guided the representation, we show the models for this example, given by the LM-entailment algorithm, and how the representation changed because of them. It is important to note that we restrict ourselves to the use of only a subset of PTL. Before determining the final representation, we begin by only allowing PTL statements of the form, $\bullet\alpha \rightarrow \beta$ or $\bullet\alpha \rightarrow \bullet\beta$, where α and β could be any formulas from the propositional logic language. The reason being that the examples we deal with can be represented reasonably with this limited language and this limiting also reduces the complexity of the analysis. Statements of the form $\alpha \to \bullet\beta$ do not carry the intuition we desire, since we do not want the properties of β, whether they are the most typical or otherwise, to apply to all α valuations. So we desire that the bullet operator be present on the antecedent side at least. Now we move on to determine which of $\bullet\alpha \to \beta$ and $\bullet\alpha \to \bullet\beta$ we are to use for representation.

8.1.1 Bullet operator on both sides?

PTL has similar semantics to the semantics of DSDL and KLM-style defeasible reasoning in that its valuations are ranked by some order of preference. Thus we begin the search for a representation method by adopting the same translation used in chapter 7. Recall from definition 5.2 that we can translate any KLM style conditional to a PTL formula as we can go from the $\alpha \mid\!\sim \beta$ to $\bullet\alpha \to \beta$ and retain the same reading. The PTL statement would be *"the most typical α are β"* which holds the same intuition as the KLM-style version, *"α typically implies β"*. This KLM-equivalent version tells us that the most typical α-valuations have to be any β-valuations and this seems reasonable. But we have already examined this representation in chapter 7 and thus wish to explore the added expressive capabilities provided by PTL but wish to do so without sacrificing an intuitive interpretation of the obligations. So the question we ask is whether having the typicality bullet operator on both sides of the implication arrow, instead of only on the antecedent side, provides a sensible reading of the obligations.

We would represent the obligations as $\bullet\alpha \to \bullet\beta$ and read them as *"the most typical α's are the most typical β's"*. This tells us that the most typical situations where one is late must be among the most typical situations where one apologises. This adds a further constraint to the obligation as we require scenarios where α holds, to be such that no other scenario where β holds should be more typical. Both versions of the obligation seem reasonable, but the representation with bullets on both sides gives us a more specific version of obligation. This suggests that having the bullets on both sides provides a sufficiently satisfactory representation for obligations and with this in mind, we proceed with this both-side representation. We initially wished to explore the added expressiveness of PTL and the reading of the obligations seems adequate to justify the use of this representation.

8.1.2 Exceptions

We look at the handling of exceptions using the current PTL representation up to this point in the chapter. During this section, we revisit the Student example introduced in section 7.3. It is an example of having a contrary-to-duty obligation with an exception and when reasoning, ideally, the exception would be more plausible than a violation. Below are the statements of the Student example: *"Students must not pay taxes"* and *"Students who work must pay taxes"*. Here are the PTL equivalents of the statements.

$$\{\bullet s \rightarrow \bullet\neg t, \bullet(s \wedge w) \rightarrow \bullet t\}.$$

The below figure shows the LM-entailment model of the statements.

2	{s,¬t,w}
1	{s,t,w},{s,t,¬w}
0	{s,¬t,¬w}

We strip the model of valuations where $\neg s$ holds, by assuming the fact s holds, as scenarios with no student are not of interest for this example. The model shows that there is an equal preference for both valuations where a student is paying taxes whether or not they are working. This is undesirable as the exceptional statement *"a student is working and therefore should pay taxes"* should be a more plausible scenario than the violation *"a student pays taxes when they are not working and therefore must not do so"*. For that to be the case, the valuation $\{s, t, \neg w\}$ would need to be a level higher than the valuation $\{s, t, w\}$, whereas they are currently on the same level of preference. This is because of the statement $\bullet(s \wedge w) \rightarrow \bullet t$ which asserts that the best valuations where there is a student who pays taxes, have to be valuations where the student is working, however, the converse need not be true. The below figure shows the model if we instead represent the statement by $\bullet(s \wedge w) \leftrightarrow \bullet t$.

$\mathscr{R}^*_{\mathcal{KB}}$	2	{s,¬t,w}, {s,t,¬w}
	1	{s,t,w}
	0	{s,¬t,¬w}

The change from the one-sided implication → to the if-and-only-if double implication ↔ provides us with a different reading. In general, if we have the non-conditional obligation of $\neg\beta$ and have that α is an exception to it, then this means that β is obligated to be the case whenever α is true. Now conversely, can we say that if the violation for the non-conditional obligation holds, that is that β holds, then the exception α is obligated to be true as well? The change to the if-and-only-if double implication, $\leftrightarrow$, in the second obligation will push the valuation $\{s, t, \neg w\}$ up a level. This causes it to be less preferred than the situation where a student who pays taxes is also a working student. So in a technical sense, this seems to be an adequate method of representing those obligations that have exceptions. The model states that the exceptions should simply be preferred to ordinary violations and this seems reasonable. The reading that if you are paying taxes then you should be a working student seems less so but it is a result of the information we have, as all that is known about taxpayers in the scenario is that they can only be a working student. Let us look at a different version of the example which generates the same result. If the statements are $\bullet s \rightarrow \bullet\neg t$, *"the most typical students are the most typical non-taxpayers"*, and $\bullet w \leftrightarrow \bullet t$, *"the most typical workers are the most typical taxpayers"*, then the model with the all valuations included is given below.

$\mathscr{R}^*_{\mathcal{KB}}$	2	{s,¬t,w}, {s,t,¬w}, {¬s,t,¬w}, {¬s,¬t,w}
	1	{s,t,w}, {¬s,t,w}
	0	{s,¬t,¬w}, {¬s,¬t,¬w}

Now if we look at the model, it states the ideal scenario where you pay taxes is one when where you are working and vice versa. This provides a more intuitive reading and if we remove the valuations where there is no student then the model will be equivalent to the initial Student example model from above. Thus the representation rule would state that we should use the double implication operator to represent contrary-to-duty obligations with exceptions. Note that this exceptions representation creates an underlying logical equivalency between statements such as "if α then β is obligatory" and "if β then α is obligatory", which represents a drawback to its use.

8.1.3 Representation summary

These are the following rules to follow to translate a deontic statement, $\bigcirc(\beta \mid \alpha)$ into a PTL statement. We would do this for all obligations within the deontic knowledge base we are translating.

1. Convert statements such as $\bigcirc(\beta \mid \alpha)$ to $\bullet\alpha \rightarrow \bullet\beta$. Note that if we have a non-conditional obligation, let's say $\bigcirc\beta$, then the antecedent will be

the tautology $\top$. So then the non-conditional obligation, $\bigcirc(\beta \mid \top)$ or more succinctly $\bigcirc(\beta)$, can be represented with $\bullet\top \to \bullet\beta$.

2. If the statement represents an exceptional contrary-to-duty obligation to another obligation let's say $\bigcirc(\beta \mid \alpha)$ then the PTL statement will become $\bullet\alpha \leftrightarrow \bullet\beta$. We state that an exceptional contrary-to-duty obligation for the obligation $\bigcirc(\beta \mid \alpha)$ would be a statement where the antecedent is not one of $\{\top, \bot, \alpha, \neg\alpha, \beta, \neg\beta\}$, and the consequent would imply $\neg\beta$.

There may also be facts in the to-be-translated knowledge base and we represent these in the usual way. With the way of representing deontic scenarios chosen, the analysis process can continue to the properties stage.

8.2 Properties

We now explore the extent to which the properties detailed in section 2 apply in our restricted PTL environment. We are not assessing whether these properties are generally satisfied by PTL for all types of statements and scenarios. In particular, we want to determine whether these properties are applicable to obligations of a similar form to those translated using the representation method we have settled upon. With there being various ways that we can represent an obligation in PTL, we will look at how we can apply each property for particular representations of obligations.
For this section, we will make use of the stronger notion of the statements being within the knowledge base, instead of just being entailed by the knowledge base, in preparation for the paradox analysis. This is due to the paradoxes' statements being part of the knowledge bases and not just entailed by them.

8.2.1 Ought Implies Can Principle

We mentioned in the discussion of this principle in chapter 7 that while working with obligations, we wish to state explicitly whether a certain fact brings up a violation. This would require the logic system used to allow a conflict, but this is not the case when we use PTL. Let's say we have a knowledge base that contains the conditionals $\bullet\top \to \bullet\alpha$ and $\bullet\top \to \bullet\neg\alpha$. There will be no valuations that satisfy the knowledge base because of the conflicting conditionals, therefore we cannot reason with this knowledge base. This implies that we have the "ought implies can" principle. In deontic form, we represent this with the derivation of $\neg \bigcirc (\alpha \land \neg\alpha)$. A benefit of this, just as with the KLM approach, is that we cannot have the previously discussed issue of deontic explosion.

Since having contradictory facts in the knowledge base causes a conflict, we will not have any facts in the knowledge base when using the LM-entailment algorithm to construct the LM-entailment model. These facts we omit from the knowledge base will be the facts not considered being background knowledge. But similarly to the previous section, we can observe which obligations a certain fact violates by checking each obligation against the fact. When there is an inconsistency and thus no valuations that satisfy the knowledge base, that indicates that we can check the obligations to determine which have been violated. After determining the violated obligations, we will then use facts after the LM-entailment algorithm constructs the ranked model. For example, if we have the knowledge base such as $\{\bullet\top \rightarrow \bullet\alpha, \bullet\top \rightarrow \bullet\beta, \neg\beta\}$, which has two obligations and a single fact, we will construct the LM-entailment model using only the subset $\{\bullet\top \rightarrow \bullet\alpha, \bullet\top \rightarrow \bullet\beta\}$. We approach dealing with exceptional obligations in the same manner. We take any background knowledge, let's say $\alpha \rightarrow \beta$ for this example, into account during the LM-entailment model construction. Then when we are checking for the derivation of an obligation with $\neg\alpha$ as the premise, let's say we are examining whether we can derive $\neg\alpha \rightarrow \beta$, we introduce the fact in the following manner. We will strip valuations from the LM-entailment model that contradict our fact and then reason with the resultant model. Thus we are now reasoning knowing that this fact has occurred and we do not want to deal with scenarios which contradict it. The intuition is that the model conveys the best-case scenario if an obligation has been violated.

8.2.2 Restricted Strengthening of the Antecedent

We present an example to illustrate the issue with having the Restricted Strengthening of the Antecedent property with our given representation. Let's say we have the following obligations: *"You should not drive under the influence"* and *"You should have a driver's license"*. Additionally, there is the fact *"You drive under the influence"*. Observe the following derivation.

$$\frac{\models d, \mathcal{KB} \mathrel{|\!\approx} \bigcirc l}{\mathcal{KB} \mathrel{|\!\approx} \bigcirc (l \mid d)}$$

Since d is a tautology we have $\top \rightarrow d$, and the derivation can be thought of as an instance of the Restricted Strengthening of the Antecedent property being applied. We include the derived obligation, *"If you drive under the influence you should have a driver's license"*, in the knowledge base and as previously mentioned, omit the fact from the knowledge base.

$$\{\bigcirc\neg d, \bigcirc(l \mid d), \bigcirc l\}$$

This set of obligations seems reasonable and intuitively consistent. The justification being that if you must do two separate things and then you violate one of them, you are still obligated to do the other. This example, rather, provides a technical issue in PTL. When we use our PTL representation, then the above example cannot be adequately modelled by:

$$\{\bullet\top \rightarrow \bullet\neg d, \bullet d \rightarrow \bullet l, \bullet\top \rightarrow \bullet l\}$$

Here we have that the most typical situations are those where both $\neg d$ and l hold. We also have that the most typical situations where d holds should be the most typical situations where l holds. The latter's reading does not seem as reasonable when taking into account the previous two. Technically, this set of PTL statements results in a model where it is not possible to have the most typical d-valuations be the most typical l-valuations as the most typical valuations of the entire model have to be not only l-valuations but also $\neg d$-valuations. This leaves us with the following model where we have scenarios where d holds being given an infinite rank thus being seen as impossible scenarios. This leaves us with the potential to derive anything if we use d as a premise, akin to deontic explosion.

$\mathscr{R}^*_{\mathcal{KB}}$	∞	$\{d, l\}, \{d, \neg l\}$
	1	$\{\neg d, \neg l\}$
	0	$\{\neg d, l\}$

This issue results from the statement $\bullet d \rightarrow \bullet l$. It is a contrary-to-duty obligation to $\bullet\top \rightarrow \bullet\neg d$ and while noting that the most typical valuations must satisfy l along with $\neg d$, we can see that the obligation $\bullet d \rightarrow \bullet l$ cannot obligate the occurrence of the most typical l situations as it violates the most typical valuations satisfying $\neg d$. This suggests there needs to be an adjustment to the Restricted Strengthening of the Antecedent property to avoid situations such as these, without disregarding the property entirely. We use the Rational Monotonicity property, previously discussed in chapter 3, to guide us in devising such a property. Once again, the property in deontic logic would look like the following.

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\beta \mid \alpha), \mathcal{KB} \mid\!\approx \neg \bigcirc (\neg\gamma \mid \alpha)}{\mathcal{KB} \mid\!\approx \bigcirc(\beta \mid \alpha \wedge \gamma)}$$

Now we revisit the example's deontic knowledge base, $\{\bigcirc\neg d, \bigcirc l\}$, which has the derived obligation $\bigcirc(l \mid d)$ removed to observe whether we can derive it with this alternative property inspired by Rational Monotonicity. The above property tells us that if we wish to derive $\bigcirc(l \mid d)$ from $\bigcirc l$ then we must not have $\bigcirc\neg d$ in the knowledge base. Since that is not the case, we get the desired result of blocking the derivation of $\bigcirc(l \mid d)$. The statement $\mathcal{KB} \mid\!\approx \neg \bigcirc (\neg\gamma \mid \alpha)$ tells us that from the knowledge base it does not follow that given α, there is an obligation to do $\neg\gamma$. This says that given α being true then γ is permitted. The PTL equivalent to this deontic formula is $\neg(\bullet\alpha \rightarrow \bullet\neg\gamma)$. We now continue to examine what the LM-entailment model tells about this property. For this property, we assume that we have a knowledge base that contains the obligation $\bigcirc(q \mid p)$, and does not contain $\bigcirc(\neg r \mid p)$, where p, q and r are propositional letters. We want to check if we can derive $\bigcirc(q \mid p \wedge\ r)$ when using the LM-entailment model.

1. Let's say we have the following knowledge base, $\{\bullet p \rightarrow \bullet q\}$. Ideally, we want to derive $\bullet(p \wedge r) \rightarrow \bullet q$ in the case where r holds and below is the LM-entailment model $\mathscr{R}^{*}_{\mathcal{KB}}$.

1	$\{p, \neg q, r\}, \{p, \neg q, \neg r\}$
0	$\{p, q, r\}, \{p, q, \neg r\}, \{\neg p, q, r\}, \{\neg p, \neg q, r\}, \{\neg p, \neg q, \neg r\}, \{\neg p, q, \neg r\}$

   When r holds, the most typical $p \wedge r$-valuation is also the most typical q-valuation thus we can derive $\bullet(p \wedge r) \rightarrow \bullet q$. This is possible since we do not have that $\bullet p \rightarrow \bullet\neg r$ (or $\bullet p \leftrightarrow \bullet\neg r$) in the knowledge base. Including this obligation would have lifted the lowest $p \wedge r$-valuation up the model and made the most typical q-valuation become a valuation where only one of p and r are true. This derivation of $\bullet(p \wedge r) \rightarrow \bullet q$ would also be blocked if we had $\bullet(p \wedge r) \rightarrow \bullet\neg q$ or $\bullet\neg(p \wedge r) \rightarrow \bullet q$ included in the knowledge base.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q\}$. Ideally, we want to derive $\bullet(p \wedge r) \rightarrow \bullet q$ in the case where r holds. The model

for this case would be similar to the model of the above case, but with the valuations that have $\neg p \wedge q$ being true, which are the $\{\neg p, q, r\}$ and $\{\neg p, q, \neg r\}$, moved up a level. The fact remains that the most typical $p \wedge r$-valuation is also the most typical q-valuation while we do not have $\bullet p \rightarrow \bullet\neg r$ (or $\bullet p \leftrightarrow \bullet\neg r$) in the knowledge base. The addition of this obligation would have the same effect as in the above case. Below is the LM-entailment model, $\mathscr{R}^{*}_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$, $\{\neg p,q, \neg r\}$, $\{\neg p,q, r\}$
0	$\{p,q, r\}$,$\{p,q, \neg r\}$, $\{\neg p,\neg q, r\}$, $\{\neg p,\neg q, \neg r\}$

We can therefore see that our alternative to the Restricted Strengthening of the Antecedent for PTL can be applied while using the LM-entailment model.

8.2.3 Weakening

$$\frac{\mathcal{KB} \mathrel{|\!\approx} \bigcirc(\beta \wedge \gamma \mid \alpha)}{\mathcal{KB} \mathrel{|\!\approx} \bigcirc(\beta \mid \alpha)}$$

For this property, we assume that we have a knowledge base that contains $\bigcirc(q \wedge r \mid p)$ where p, q and r are propositional letters. We want to check if $\bigcirc(q \mid p)$ can be derived when using the LM-entailment model.

1. Let's say we have the following knowledge base, $\{\bullet p \rightarrow \bullet(q \wedge r)\}$. Ideally, we want to derive $\bullet p \rightarrow \bullet q$. Below is the LM-entailment model, $\mathscr{R}^{*}_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$, $\{p,q, \neg r\}$
0	$\{p,q, r\}$, $\{\neg p,q, r\}$, $\{\neg p,\neg q, r\}$, $\{\neg p,\neg q, \neg r\}$, $\{\neg p,q, \neg r\}$

The most typical p-valuation is the most typical q-valuation and is also the most typical r-valuation therefore we can derive both $\bullet p \rightarrow \bullet q$ and $\bullet p \rightarrow \bullet r$. Note that this will not always be the case as we can have, for example, that the best q-valuation be a $\neg r$-valuation and vice versa for the best r-valuation. This would mean the best $p \wedge q$-valuation would not be the best q-valuation or the best r-valuation.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet(q \wedge r)\}$. Ideally, we want to derive $\bullet p \rightarrow \bullet q$. This model is the same as the above case except the following valuation, $\{\neg p, q, r\}$, is pushed up a level. So we can still derive both $\bullet p \rightarrow \bullet q$ and $\bullet p \rightarrow \bullet r$. This same note on applicability as above applies here. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p, \neg q, r\}$, $\{p, \neg q, \neg r\}$, $\{p, q, \neg r\}$, $\{\neg p, q, r\}$
0	$\{p, q, r\}$, $\{\neg p, \neg q, r\}$, $\{\neg p, \neg q, \neg r\}$, $\{\neg p, q, \neg r\}$

The Weakening property is applicable to both ordinary obligations and those that represent exceptions, under certain circumstances. The property does not hold generally as it is not always the case that $\bullet(q \wedge r) \rightarrow \bullet q$.

8.2.4 Deontic Detachment

$$\frac{\mathcal{KB} \mathrel{|\!\!\approx} \bigcirc(\beta \mid \alpha), \mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\alpha}{\mathcal{KB} \mathrel{|\!\!\approx} \bigcirc\beta}$$

For this property, we assume that we have a knowledge base that contains at least $\bigcirc(q \mid p)$ and $\bigcirc p$ where p, q and r are propositional letters. We want to check the derivability of $\bigcirc q$ when using the LM-entailment model. There are only two cases to consider as the case with two non-conditional obligations is trivial.

1. Let's say we have the following knowledge base, $\{\bullet p \rightarrow \bullet q, \bullet\top \rightarrow \bullet p\}$. Ideally, we want to derive $\bullet\top \rightarrow \bullet q$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p, \neg q\}$, $\{\neg p, \neg q\}$, $\{\neg p, q\}$
0	$\{p, q\}$

The most typical valuation is $\{p, q\}$ so we can derive $\bullet\top \rightarrow \bullet q$.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet\top \rightarrow \bullet p\}$. Ideally, we want to derive $\bullet\top \rightarrow \bullet q$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p, \neg q\}$, $\{\neg p, \neg q\}$, $\{\neg p, q\}$
0	$\{p, q\}$

As in the above case, the most typical valuation is $\{p, q\}$ thus $\bullet\top \to \bullet q$ can be derived.

The Deontic Detachment property is applicable to both scenarios where we have a non-conditional obligation and another obligation, ordinary or exceptional.

8.2.5 Factual Detachment

$$\frac{\mathcal{KB} \mathrel{|\!\approx} \bigcirc(\beta \mid \alpha), \models \alpha}{\mathcal{KB} \mathrel{|\!\approx} \bigcirc\beta}$$

For this property, we assume that we have a knowledge base that contains at least $\bigcirc(q \mid p)$ and the fact p where p, q and r are propositional letters. We want to check the derivability of $\bigcirc q$ when using the LM-entailment model. There are only three cases to look at as the non-conditional obligation check is trivial.

1. Let's say we have the following knowledge base, $\{\bullet p \to \bullet q\}$. Ideally, we want to derive $\bullet\top \to \bullet q$ in the case where p holds. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q\}$
0	$\{p,q\}$, $\{\neg p,q\}$, $\{\neg p,\neg q\}$

   When p is true, the most typical valuation is $\{p, q\}$ therefore the desired derivation of $\bullet\top \to \bullet q$ holds.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q\}$. Ideally, we want to derive $\bullet\top \to \bullet q$ in the case where p holds. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q\}$,$\{\neg p,q\}$
0	$\{p,q\}$, $\{\neg p,\neg q\}$

   When we have p, the most typical valuation is $\{p, q\}$ and we can derive $\bullet\top \to \bullet q$.

Given a fact p, we can see that Factual Detachment is applicable to both types of obligations we deal with.

8.2.6 Distribution

$$\text{If } \models \alpha \rightarrow \beta, \text{ then } \mathcal{KB} \mathrel{|\!\approx} \bigcirc\alpha \rightarrow \bigcirc\beta$$

Recall that this tells us that if we have a fact which states α implies β, if we have that one is obligated to do α then one is also obligated to do β. So using the chosen PTL representation for this property, we assume we have $\alpha \rightarrow \beta$ as a tautology and the non-conditional obligation $\bullet\top \rightarrow \bullet\alpha$ in the knowledge base. The question is whether we can derive $\bullet\top \rightarrow \bullet\beta$. We have the following knowledge base, $\{\bullet\top \rightarrow \bullet p, p \rightarrow q\}$ where p and q are propositional letters. Ideally, we want to derive $\bullet\top \rightarrow \bullet q$. The fact $p \rightarrow q$ means we cannot have the valuation $\{p, \neg q\}$ in the LM-entailment model when we assume the fact to hold. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

$\mathscr{R}^*_{\mathcal{KB}}$:

1	$\{\neg p, q\}$, $\{\neg p, \neg q\}$
0	$\{p, q\}$

We can derive $\bullet\top \rightarrow \bullet q$, as the most typical valuation in the model is $\{p, q\}$ and we therefore have the Distribution property with our given representation.

8.2.7 Conjunction

$$\frac{\mathcal{KB} \mathrel{|\!\approx} \bigcirc(\beta \mid \alpha), \mathcal{KB} \mathrel{|\!\approx} \bigcirc(\gamma \mid \alpha)}{\mathcal{KB} \mathrel{|\!\approx} \bigcirc(\beta \wedge \gamma \mid \alpha)}$$

For this property, we assume that we have a knowledge base that contains at least $\bigcirc(q \mid p)$ and $\bigcirc(r \mid p)$ where p, q and r are propositional letters. We want to check the derivability of $\bigcirc(q \wedge r \mid p)$ when using the LM-entailment model.

1. Let's say we have the following knowledge base, $\{\bullet p \rightarrow \bullet q, \bullet p \rightarrow \bullet r\}$. Ideally, we want to derive $\bullet p \rightarrow \bullet(q \wedge r)$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p, \neg q, r\}$, $\{p, \neg q, \neg r\}$, $\{p, q, \neg r\}$
0	$\{p, q, r\}$, $\{\neg p, q, r\}$, $\{\neg p, \neg q, r\}$, $\{\neg p, \neg q, \neg r\}$, $\{\neg p, q, \neg r\}$

   We can see that the most typical p-valuation is $\{p, q, r\}$ so we can derive $\bullet p \rightarrow \bullet(q \wedge r)$.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet p \rightarrow \bullet r\}$. Ideally, we want to derive $\bullet p \rightarrow \bullet(q \wedge r)$. Below is the LM-entailment model, $\mathscr{R}^{*}_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$, $\{p,q, \neg r\}$,$\{\neg p,q, \neg r\}$,$\{\neg p,q, r\}$
0	$\{p,q, r\}$, $\{\neg p,\neg q, r\}$, $\{\neg p,\neg q, \neg r\}$

   Again. $\{p, q, r\}$ is the most typical p-valuation, so we can derive $\bullet p \rightarrow \bullet(q \wedge r)$.

3. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet p \leftrightarrow \bullet r\}$. Ideally, we want to derive $\bullet p \rightarrow \bullet(q \wedge r)$. Below is the LM-entailment model, $\mathscr{R}^{*}_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$, $\{p,q, \neg r\}$,$\{\neg p,q, \neg r\}$,$\{\neg p,q, r\}$,$\{\neg p,\neg q, r\}$
0	$\{p,q, r\}$, $\{\neg p,\neg q, \neg r\}$

   $\{p, q, r\}$ is the most typical p-valuation so we can derive $\bullet p \rightarrow \bullet(q \wedge r)$.

The Conjunction property is applicable to the various combinations of our obligation types.

8.2.8 Disjunction

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\beta \mid \gamma), \mathcal{KB} \mid\!\approx \bigcirc(\beta \mid \alpha)}{\mathcal{KB} \mid\!\approx \bigcirc(\beta \mid \alpha \vee \gamma)}$$

For this property, we assume that we have a knowledge base that contains at least $\bigcirc(q \mid p)$ and $\bigcirc(r \mid p)$ where p, q and r are propositional letters. We want to check the derivability of $\bigcirc(q \mid r \vee p)$ when using the LM-entailment model.

1. Let's say we have the following knowledge base, $\{\bullet p \rightarrow \bullet q, \bullet r \rightarrow \bullet q\}$. Ideally, we want to derive $\bullet(p \vee r) \rightarrow \bullet q$. Below is the LM-entailment model, $\mathscr{R}^{*}_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$,$\{\neg p,\neg q, r\}$
0	$\{p,q, r\}$, $\{\neg p,q, r\}$, $\{\neg p,\neg q, \neg r\}$, $\{\neg p,q, \neg r\}$,$\{p,q, \neg r\}$

The most typical p-valuations are also the most typical q-valuations. We also have $\{p, q, r\}$ and $\{\neg p, q, r\}$ as the most typical r-valuations therefore $\bullet(p \vee r) \rightarrow \bullet q$ holds.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet r \rightarrow \bullet q\}$. Ideally, we want to derive $\bullet(p \vee r) \rightarrow \bullet q$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$,$\{\neg p,\neg q, r\}$,$\{\neg p,q, r\}$,$\{\neg p,q, \neg r\}$
0	$\{p,q, r\}$, $\{\neg p,\neg q, \neg r\}$,$\{p,q, \neg r\}$

We have $\{p, q, r\}$ and $\{p, q, \neg r\}$ as the most typical p-valuations with the former being the most typical r-valuation. Both are the most typical q-valuation therefore $\bullet(p \vee r) \rightarrow \bullet q$ holds.

3. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet r \leftrightarrow \bullet q\}$. Ideally, we want to derive $\bullet(p \vee r) \rightarrow \bullet q$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q, r\}$,$\{p,\neg q, \neg r\}$,$\{\neg p,\neg q, r\}$,$\{\neg p,q, r\}$,$\{\neg p,q, \neg r\}$,$\{p,q, \neg r\}$
0	$\{p,q, r\}$, $\{\neg p,\neg q, \neg r\}$

The most typical p-valuation and most typical r-valuation is $\{p, q, r\}$ so $\bullet(p \vee r) \rightarrow \bullet q$ holds.

Similarly to the Conjunction property, given our combinations of obligation types, the Disjunction property is one we can still apply.

8.2.9 Transitivity

$$\frac{\mathcal{KB} \mid\!\approx \bigcirc(\gamma \mid \beta), \mathcal{KB} \mid\!\approx \bigcirc(\beta \mid \alpha)}{\mathcal{KB} \mid\!\approx \bigcirc(\gamma \mid \alpha)}$$

For this property, we assume that we have a knowledge base that contains at least $\bigcirc(q \mid p)$ and $\bigcirc(r \mid q)$ where p, q and r are propositional letters. We want to check the derivability of $\bigcirc(r \mid p)$ when using the LM-entailment model.

1. Let's say we have the following knowledge base, $\{\bullet p \rightarrow \bullet q, \bullet q \rightarrow \bullet r\}$. Ideally, we want to derive $\bullet p \rightarrow \bullet r$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q,\ r\},\{p,\neg q,\ \neg r\},\{p,q,\ \neg r\},\{\neg p,q,\ \neg r\}$
0	$\{p,q,\ r\},\ \{\neg p,q,\ r\},\ \{\neg p,\neg q,\ \neg r\},\{\neg p,\neg q,\ r\}$

We have the most typical p-valuation being the most typical r-valuation so $\bullet p \to \bullet r$ can be derived.

2. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet q \to \bullet r\}$. Ideally, we want to derive $\bullet p \to \bullet r$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q,\ r\},\{p,\neg q,\ \neg r\},\{\neg p,q,\ r\},\{\neg p,q,\ \neg r\},\{p,q,\ \neg r\}$
0	$\{p,q,\ r\},\ \{\neg p,\neg q,\ \neg r\},\{\neg p,\neg q,\ r\}$

   $\{p, q, r\}$ is the most typical p-valuation so we can derive $\bullet p \to \bullet r$.

3. Let's say we have the following knowledge base, $\{\bullet p \leftrightarrow \bullet q, \bullet q \leftrightarrow \bullet r\}$. Ideally, we want to derive $\bullet p \to \bullet r$. Below is the LM-entailment model, $\mathscr{R}^*_{\mathcal{KB}}$.

1	$\{p,\neg q,\ r\},\{p,\neg q,\ \neg r\},\{\neg p,\neg q,\ r\},\{\neg p,q,\ r\},\{\neg p,q,\ \neg r\},\{p,q,\ \neg r\}$
0	$\{p,q,\ r\},\ \{\neg p,\neg q,\ \neg r\}$

   We can derive $\bullet p \to \bullet r$ since the most typical p-valuation is the most typical r-valuation.

Given the combinations of obligation types, we have that the Transitivity property can be derived within the LM-entailment model.

8.2.10 Conclusions

We observe that we can apply most of the properties within PTL under specific scenarios. Instead of having the Restricted Strengthening of the Antecedent property with our given representation, we use an alternative property inspired by the Rational Monotonicity property of the KLM approach. With the general satisfaction of the properties within PTL still to be checked, analysis of the properties that are held within the underlying logic could also be valuable. We proceed to examine how the properties interact in the LM-entailment models of the paradoxes.

8.3 Paradoxes

The following section details the results of reasoning about the paradoxes using LM-entailment. The LM-entailment models are constructed using the LM-entailment algorithm on knowledge bases translated to the PTL language with the chosen PTL representation method. A reminder that the desired interpretation for the paradoxes, since we have "ought implies can" principle, is for conflicts to be unwanted. We rather aim to find the obligation to act as best as possible once a violation has been committed [21, 23].

8.3.1 Forrester's paradox

Recall that this paradox comprises the following three statements: *"You must not kill anybody"*, *"If you kill someone then you must kill them gently"* and *"You killed someone "*. There is also the background knowledge that states that *"Killing gently implies killing"*. Along with the following set of PTL statements comes the translated version of the background knowledge, $g \rightarrow k$, and the fact k. We assume the background knowledge to hold throughout while we consider the fact to hold when we attempt to determine actual obligations.

$$\{\bullet\top \rightarrow \bullet\neg k, \bullet k \rightarrow \bullet g\}$$

The violated obligation is $\bullet\top \rightarrow \bullet\neg k$ as no valuation could satisfy a knowledge base which has both the fact k and this obligation. The LM-entailment algorithm constructs the following model of valuations.

$\mathscr{R}^*_{\mathcal{KB}}$	2	$\{\neg g, k\}$
	1	$\{g, k\}$
	0	$\{\neg g, \neg k\}$

From the background knowledge $g \rightarrow k$, we have that the valuation $\{g, \neg k\}$ cannot be in the model. The model tells us that the most typical of all scenarios is when there has been no killing, gently or otherwise. The next most plausible scenario is when one has killed gently and the least plausible is where one killed but did not to do so gently. We now present the evaluation of the paradox's problematic derivations.

Restricted Strengthening of the Antecedent, Weakening and Conjunction

1.	$\bigcirc\neg k \rightarrow_{W} \bigcirc\neg g$
2.	$\bigcirc\neg g \rightarrow_{RSA} \bigcirc(\neg g \mid k)$
3.	$\{\bigcirc(\neg g \mid k), \bigcirc(g \mid k)\} \rightarrow_{Conj} \bigcirc(\neg g \wedge g \mid k)$

The Weakening property allows for the derivation of $\bigcirc\neg g$ from $\bigcirc\neg k$ which is reasonable. Working with the LM-entailment model $\mathscr{R}^*_{\mathcal{KB}}$, we can go from $\bullet\top \rightarrow \bullet\neg k$ to $\bullet\top \rightarrow \bullet\neg g$ as the model shows that the most typical valuation is not only a $\neg k$-valuation but also satisfies $\neg g$. This mimics the deontic derivation using the Weakening property. But unlike in the above derivations, one cannot derive $\bigcirc(\neg g \mid k)$ from using Restricted Strengthening of the Antecedent. Using our PTL alternative to Restricted Strengthening of the Antecedent, the presence of the statement $\bullet\top \rightarrow \bullet\neg k$ blocks $\bullet k \rightarrow \bullet\neg g$ from being derived. We can see this in the model as the best k valuations are g valuations in this model and thus cannot be a $\neg g$-valuation.

Factual Detachment and Conjunction

1.	$\{\bigcirc(g \mid k), k\} \rightarrow_{FD} \bigcirc g$
2.	$\{\bigcirc\neg k, \bigcirc g\} \rightarrow_{Conj} \bigcirc(\neg k \wedge g)$

The first derivation where the Factual Detachment property allows for the derivation of $\bigcirc g$ from the pair $\bigcirc(g \mid k)$ and the fact k, is also possible in the LM-entailment model. When we assume k, the valuations where $\neg k$ holds are stripped from the model which leaves us with the following model. Here we get $\bullet\top \rightarrow \bullet g$ as an actual obligation, which represents the obligation $\bigcirc g$, since the most typical of the two valuations is one where g holds.

$\mathscr{R}^*_{\mathcal{KB}}$		
	1	$\{\neg g, k\}$
	0	$\{g, k\}$

This model shows us that once we have assumed that a killing has occurred that the derivation of $\bullet\top \rightarrow \bullet\neg k$ is blocked, and thus the problematic derivation of $\bullet\top \rightarrow \bullet(\neg k \wedge g)$ is avoided when we assume k.

8.3.2 Chisholm's paradox

Recall that this paradox comprises the following four statements: *"Jones must go assist his neighbour"*, *"If Jones goes to assist his neighbour then he* *must tell them that he is coming"*, *"If Jones does not go to assist his neighbour then he must not tell them that he is coming"* and *"Jones does not go to assist his neighbour"*. Along with the following set of PTL statements comes the fact $\neg a$.

$$\{\bullet\top \rightarrow \bullet a, \bullet a \rightarrow \bullet t, \bullet\neg a \rightarrow \bullet\neg t\}$$

The LM-entailment algorithm constructs the following model of valuations. The violated obligation is $\bullet\top \rightarrow \bullet a$ as no valuation could satisfy a knowledge base which has both the fact $\neg a$ and this obligation.

$\mathscr{R}^*_{\mathcal{KB}}$

2	{¬a , t}
1	{¬a , ¬t},{a, ¬t}
0	{a , t}

We observe that the most typical situation is where Jones goes to assist and tells his neighbour that he is coming to assist. At a less typical level, we have the scenario where Jones does not go assist and does not tell his neighbour as well as the scenario where Jones goes to assist but does not tell his neighbour. The least plausible scenario is when Jones does not go assist yet tells his neighbour that he is coming to assist. Note that the obligation $\bullet a \rightarrow \bullet t$ is no longer necessary and is implied by the other two. This is due to the most typical a valuation automatically becoming the best t valuation from $\bullet\top \rightarrow \bullet a$ and $\bullet\neg a \rightarrow \bullet\neg t$. This is certainly an undesirable scenario given by the representation we use. The obligation for Jones to go, and the obligation which states how he should act if he does not go, should have no bearing how he should act if he goes to assist [5].

Deontic Detachment, Factual Detachment and Conjunction

1.	$\{\bigcirc(t \mid a), \bigcirc a\} \rightarrow_{DD} \bigcirc t$
2.	$\{\bigcirc(\neg t \mid \neg a), \neg a\} \rightarrow_{FD} \bigcirc\neg t$
3.	$\{\bigcirc t, \bigcirc\neg t\} \rightarrow_{Conj} \bigcirc t \wedge \neg t$

When we take $\neg a$, *"Jones does not go to assist his neighbour"*, as a fact, we get the following model with the valuations where a holds being stripped away.

$\mathscr{R}^*_{\mathcal{KB}}$

1	{¬a , t}
0	{¬a , ¬t}

We can still derive $\bigcirc\neg t$, $\bullet\top \rightarrow \bullet\neg t$, such as with Factual Detachment but the Deontic Detachment derivation of $\bigcirc t$, $\bullet\top \rightarrow \bullet t$, is blocked. This is because of the two obligations $\bullet\top \rightarrow \bullet a$ and $\bullet a \rightarrow \bullet t$, which say *"Jones must go assist his neighbour"* and *"If Jones goes to assist his neighbour then he must tell them that he is coming"* respectively, not being derivable.

Transitivity and Restricted Strengthening of the Antecedent

1.	$\{\bigcirc(t \mid a), \bigcirc a\} \rightarrow_{Trans} \bigcirc t$
2.	$\bigcirc t \rightarrow_{RSA} \bigcirc(t \mid \neg a)$
3.	$\{\bigcirc(t \mid \neg a), \bigcirc(\neg t \mid \neg a)\} \rightarrow_{Conj} \bigcirc(t \wedge \neg t \mid \neg a)$

We can use the Transitivity property to go from the two obligations $\bigcirc(t \mid a)$ and $\bigcirc a$ to the non-conditional obligation $\bigcirc t$. And we can still perform this derivation as we can derive $\bullet\top \rightarrow \bullet t$ since $\{a, t\}$ is the most typical valuation in the model. But the derivation of $\bigcirc(t \mid \neg a)$ from $\bigcirc t$ via the Restricted Strengthening of the Antecedent property is blocked while using the alternative property because we have $\bigcirc a$ in the knowledge base. The obligation $\bigcirc(t \mid \neg a)$ would be $\bullet\neg a \rightarrow \bullet t$ in PTL, which is blocked in the model as the most typical valuation that satisfies $\neg a$ is a valuation where $\neg t$ holds.

8.3.3 Fence paradox

Recall that this paradox comprises the following four statements: *"There must not be a fence"*, *"If there is a fence then it must be a white fence"*, *"If there is a dog then there must be a white fence"* and *"There is a fence"*. Along with the following set of PTL statements comes the fact f.

$$\{\bullet\top \rightarrow \bullet\neg f, \bullet f \rightarrow \bullet(f \wedge w), \bullet d \leftrightarrow \bullet(f \wedge w)\}$$

The LM-entailment algorithm constructs the following model of valuations. The violated obligation is $\bullet\top \rightarrow \bullet\neg f$, as no valuation could satisfy a knowledge base which has both this obligation and the fact f.

$\mathscr{R}^*_{\mathcal{KB}}$	2	{¬d,f,w},{d,f,¬w}, {d,¬f,¬w},{d,¬f,w}, {¬d,f,¬w}
	1	{d,f,w}
	0	{¬d,¬f,w}, {¬d,¬f,¬w}

Note that this paradox is analogous to the Student example in section 8.1.2. The valuations where there is not a dog or a fence are the most typical. On the next level, we have the exceptional case of a dog being present and therefore we require a white fence to be present as well. Observe that the valuation where there is a white fence with no dog has a rank of two and is one of the least typical valuations. When we take the fact f to be true, which tells us that there is a fence present, we get the following model.

$\mathscr{R}^*_{\mathcal{KB}}$	1	{¬d,f,w},{d,f,¬w} {¬d,f,¬w}
	0	{d,f,w}

We can see that the least typical valuations are the ones where there is either a white fence with no dog present, or a fence which is not white and the presence of a dog is irrelevant. The wanted derivation in this example would be to have $\bigcirc(f \wedge w \mid d)$ be considered more typical than $\bigcirc(f \wedge w \mid \neg d)$ and this can be seen in the model. This tells us that when there is a violation of the non-conditional obligation to not have a fence, that it is because we have a scenario where the exception, there being a dog, has occurred. Recall that we wanted the presence of a white fence, meaning that $f \wedge w$ holds, to imply that there is a dog present, d also holds. In PTL we can derive $\bullet(f \wedge w) \rightarrow \bullet d$, due to our translation of the exceptional obligations, which is what we initially desired.

8.3.4 Trump/Kim paradox

Recall that this paradox comprises the following four statements: *“Trump must not be told the secret”*, *“Kim must not be told the secret”*, *“If you tell Trump then you must tell Kim“* and *”If you tell Kim then you must tell Trump”*. We translate the obligations into the following set of PTL statements. We are treating the latter two obligations as exceptions and therefore use the $\leftrightarrow$ representation. This means we only need to have one of these obligations in the knowledge base.

$$\{\bullet\top \rightarrow \bullet\neg t, \bullet\top \rightarrow \bullet\neg k, \bullet t \leftrightarrow \bullet k\}$$

The LM-entailment algorithm constructs the following model of valuations. The most typical scenario is when neither of the men are told the secret and the next most typical states that they should both be told the secret. The least plausible situations have only one of Trump and Kim being told the secret.

$\mathscr{R}^*_{\mathcal{KB}}$	2	$\{t, \neg k\}, \{\neg t, k\}$
	1	$\{t, k\}$
	0	$\{\neg t, \neg k\}$

Restricted Strengthening of the Antecedent, Weakening and Conjunction

1.	$\{\bigcirc\neg t, \bigcirc\neg k\} \rightarrow_{Conj} \bigcirc(\neg t \wedge \neg k)$
2.	$\bigcirc(\neg t \wedge \neg k) \rightarrow_{W} \bigcirc\neg(t \wedge k)$
3.	$\bigcirc\neg(t \wedge k) \rightarrow_{RSA} \bigcirc(\neg(t \wedge k) \mid t)$
4.	$\{\bigcirc(\neg(t \wedge k) \mid t), \bigcirc(t \wedge k \mid t\} \rightarrow_{Conj} \bigcirc(\neg(t \wedge k) \wedge t \wedge k \mid t)$

We can derive $\bullet\top \rightarrow \bullet(\neg t \wedge \neg k)$ from the above LM-entailment model as the most typical valuation in the model is $\{\neg t, \neg k\}$. This represents the first derivation using Conjunction in the above table. We can then derive $\bullet\top \rightarrow \bullet\neg(t \wedge k)$, which is the application of the Weakening property in the table. The derivation via Restricted Strengthening of the Antecedent is blocked by our alternative in this case since $\bullet\top \rightarrow \neg \bullet t$ is in our knowledge base. When we take t as true, we have the following model:

$\mathscr{R}^*_{\mathcal{KB}}$	1	$\{t, \neg k\}$
	0	$\{t, k\}$

When t is true then one cannot derive $\bullet t \rightarrow \bullet\neg(t \wedge k)$ since the most typical valuation where t is true is a $t \wedge k$-valuation.

8.3.5 Van Fraassen's paradox

Recall that this paradox comprises the following two statements: *"You must honour your father or honour your mother"* and *"You must not honour your mother"*. The translated obligations are in the following PTL set.

$$\{\bullet\top \rightarrow \bullet(f \vee m), \bullet\top \rightarrow \bullet\neg m\}$$

The LM-entailment algorithm constructs the following model of valuations. Of the four valuations, the most typical tells us that only the father has been honoured while the remaining cases are less typical.

1	$\{f, m\}, \{\neg f, m\}, \{\neg f, \neg m\}$
0	$\{f, \neg m\}$

Weakening and Conjunction

1.	$\{\bigcirc(f \vee m),\ \bigcirc\neg m\} \rightarrow_{Conj} \bigcirc(f \wedge \neg m)$
2.	$\bigcirc(f \wedge \neg m) \rightarrow_{W} \bigcirc f$

The paradox's problematic derivation involves the Weakening and Conjunction properties. We can derive the desired obligation $\bigcirc f$, which is $\bullet\top \rightarrow \bullet f$ in PTL, from this model as the most typical of all the valuations is $\{f, \neg m\}$. Note that we cannot introduce the obligation $\bullet\top \rightarrow \bullet m$ otherwise there would be a conflict, hence the deontic explosion issue is blocked.

In the alphabetic variant of Van Fraassen's paradox, the Apples-and Pears example, we have the following PTL obligation set which translates the two obligations *"You must buy apples or buy pears"* and *"You mustn't buy apples"*.

$$\{\bullet\top \rightarrow \bullet(a \vee p), \bullet\top \rightarrow \bullet\neg a\}$$

The LM-entailment algorithm constructs the following model of valuations. Much like the Van Fraassen example, the model tells us that the most typical scenario is when pears have been bought with no apples bought. The remaining valuations are all less plausible.

$\mathscr{R}^{*}_{\mathcal{KB}}$	1	$\{p, a\}, \{\neg p, a\}, \{\neg p, \neg a\}$
	0	$\{p, \neg a\}$

We give this variant's problematic derivation, detailed in section 4.4.

1.	$\{\bigcirc a \vee p, \bigcirc\neg a\} \rightarrow_{Conj} \bigcirc\neg a \wedge p$
2.	$\bigcirc\neg a \wedge p \rightarrow_{W} \bigcirc p$
3.	$\bigcirc p \rightarrow_{RSA} \bigcirc(p \mid a)$

We can derive the desired obligation $\bullet\top \rightarrow \bullet p$ but the derivation of the conditional obligation $\bullet a \rightarrow \bullet p$, through Restricted Strengthening of the Antecedent, is blocked. This derivation is blocked because the most typical valuations where a is true, $\{p, a\}$ and $\{\neg p, a\}$, are on the same level but only one satisfies p. Besides this, there is a valuation where p is true which is more typical than both $\{p, a\}$ and $\{\neg p, a\}$, with that valuation being $\{p, \neg a\}$.

8.4 Conclusion

Similarly to the KLM approach, we have an intuitive method of representation with the use of PTL's bullet operator. We translate an obligation such as $\bigcirc(\beta \mid \alpha)$ into $\bullet\alpha \rightarrow \bullet\beta$. We also have a representation for exceptions with the use of the double implication, $\leftrightarrow$, to modify a translated PTL obligation into an exceptional obligation. We represent exceptional obligations with PTL formulas such as $\bullet\alpha \leftrightarrow \bullet\beta$. And we show the deontic properties to be reasonably applicable under certain conditions except for Restricted Strengthening of the Antecedent, which causes an issue with our given representation. But the adoption of a property similar to the KLM approach's Rational Monotonicity gave a viable alternative to Restricted Strengthening of the Antecedent for use during the analysis of the paradoxes. Although the properties are fairly applicable, there is the lack of general satisfaction in PTL. Thus it is important to use the PTL representation carefully. Similar to the handling of the paradoxes with the KLM approach, the undesirable derivations are blocked while using LM-entailment. This also includes the issues given in the Fence paradox, as there is a way to represent the exceptions explicitly. Despite having the "ought implies can" principle, we can detect violations, with a similar method to that used in chapter 7, using a consistency check for valuations which satisfy the knowledge base's statements.

Chapter 9

Deontic Analysis Using PT-entailment

In this chapter, we study the use of the PT-entailment method in a deontic environment. Recall that the PT-entailment method involves the construction of models in addition to the single LM-entailment model if this is possible. Similarly to the previous chapter, we first determine how to represent the deontic statements. Then we detail how the properties which we checked in the previous chapter also apply for PT-entailment. Thereafter, we analyse specially constructed deontic examples since we cannot analyse the paradoxes for reasons detailed in section 9.3.

9.1 Representation

Since we are using the same logic language, that of PTL, we will retain the same representation from the LM-entailment chapter since it has an intuitive reading which we are satisfied with. Therefore, we use the same representation method from the LM-entailment chapter during the analysis in this chapter. Thus we translate obligations such as $\bigcirc(\beta \mid \alpha)$ into PTL formulas such as $\bullet\alpha \to \bullet\beta$. The representation of exceptions using the double implication arrow, $\leftrightarrow$, is also carried over from the LM-entailment chapter. So we represent exceptional obligations using statements such as $\bullet\alpha \leftrightarrow \bullet\beta$. Facts also keep their usual representation.

9.2 Properties

The LM-entailment models, which we used to investigate the applicability of the properties in the previous chapter, have all their valuations being as low as possible in the models. In other words, the valuations are as typical as possible, which signifies that we cannot produce additional models for the PT-entailment method. We can then use those single LM-entailment models to check the properties. For the sake of brevity, we avoid reassessing these LM-entailment models and instead carry over the properties from chapter 8 to this PT-entailment analysis. Intuitively, we state that we can apply these properties in deontic scenarios in a similar manner for both LM-entailment and PT-entailment.

9.3 Paradoxes

Ideally, we would continue this analysis with the already-presented paradoxes as we have established how to interpret their shortcomings and ideal solutions. However, when dealing with the paradoxes, the LM-entailment models rank the valuations as typically as possible. This means that the set of minimal models produced by the PT-entailment method will only contain the single LM-entailment model. Therefore, our PT-entailment derivations would be identical to that of LM-entailment for the paradoxes. Because of this, we constructed examples that are meant to be both representative of deontic intuition and facilitate the production of multiple models, in addition to the single LM-entailment model, by the PT-entailment method. The results of these examples can then demonstrate the differences between the derivations of LM-entailment and PT-entailment. For the detection of violations, we apply the same method from chapter 8 when we are presented with a fact. We then assess how PT-entailment handles these novel problems.

9.3.1 Contrary-to-duty example

The first example is a simple contrary-to-duty scenario. The PTL knowledge base we have is $\{\bullet\top \rightarrow \bullet\neg l, \bullet l \rightarrow \bullet a\}$. We can read these obligations as *"You should not be late for work"* and *"If you are late then you must apologise"*. The PT-entailment models are now given with the ∞-ranked valuations omitted and $\mathscr{R}_1$ being the LM-entailment model.

$\mathscr{R}_1$	0	$\{\neg l, \neg a\}$, $\{\neg l, a\}$

$\mathscr{R}_2$	2	$\{l, \neg a\}$
	1	$\{l, a\}$,$\{\neg l, a\}$
	0	$\{\neg l, \neg a\}$

The most problematic LM-entailment derivation, via the model $\mathscr{R}_1$, is that $\neg l$ is a tautology. This is a fact that tells us that *"You are not late"* and this is clearly too strong of a derivation considering the obligations we have. But with PT-entailment, this fact cannot be derived since $\mathscr{R}_2$ does not entail $\neg l$ being a tautology as there are valuations that satisfy l. There is another issue from $\mathscr{R}_1$ and that is the ability to derive anything when we have $\bullet l$ as the antecedent. The reason being that there are no valuations that satisfy l, which allows us to conclude whatever obligation we wish if we use l as our premise. This is akin to the undesirable deontic explosion principle discussed in various parts of the dissertation. Recall this when a violation allows one to derive any obligation from a knowledge base. The conditional obligation $\bullet l \to \bullet a$, which we deem desirable and states that *"if you are late then you must apologise"*, from PT-entailment while blocking the explosion when $\bullet l$ is the antecedent. Using $\mathscr{R}_1$, we can also derive $\bullet a \to \bullet\neg l$ which tells us that *"If you apologise then you must not be late to work"* which certainly does not align with our intuition. If there is a situation where one has to apologise then ideally they would be a latecomer. PT-entailment blocks this as the most typical valuations where a is true are $\{l, a\}$ and $\{\neg l, a\}$ so we can only derive $\bullet a \to \bullet(\neg l \vee l)$ which is more reasonable.

9.3.2 Extended contrary-to-duty example

We proceed with contrary-to-duty example that we consider an extension of the standard form with two obligations. This example has additional complexity in the linking of the contrary-to-duty obligations to the non-conditional obligation. Let's say we have the knowledge base $\{\bullet\top \to \bullet(s \wedge \neg p), \bullet\neg s \to \bullet\neg f, \bullet p \to \bullet f\}$. We can read these obligations as *"You should have a student card and should not park on campus"*, *"If you do not have a student card then you should not pay a fine"* and *"If you park on campus then you should pay a fine"*. This example has two contrary-to-duty statements which have contradictory consequents. The PT-entailment models are now given with the ∞-ranked valuations omitted and $\mathscr{R}_1$ being the LM-entailment model.

$\mathscr{R}_1$	0	$\{\neg f, s, \neg p\}, \{f, s, \neg p\}$

$\mathscr{R}_2$	2.	$\{f, \neg s, \neg p\}$
	1.	$\{\neg f, s, \neg p\}, \{\neg f, \neg s, \neg p\}$
	0.	$\{f, s, \neg p\}$

$\mathscr{R}_3$	2.	$\{\neg f, s, p\}$
	1.	$\{f, s, p\}, \{f, s, \neg p\}$
	0.	$\{\neg f, s, \neg p\}$

We can derive both s and $\neg p$ as tautologies from the LM-entailment model, $\mathscr{R}_1$, which are facts that read as *"You have a student card"* and *"You did not park"*. These are further instances of LM-entailment producing too strong a derivation in a contrary-to-duty scenario. And we can see that even in this extended example, the additional PT-entailment models block the derivations of facts from the sole LM-entailment model. Then similarly to the contrary-to-duty example in the previous section, if we have $\bullet\neg p$ or $\bullet s$ as the antecedent, we could derive any conditional obligation via deontic explosion. The presence of valuations where p and $\neg s$ both hold within the additional PT-entailment models prevents the deontic explosion of obligations. In $\mathscr{R}_1$, we can derive $\bullet\neg f \to \bullet s$ which reads as *"If you do not pay a fine then you should have a student card"*. This reading is not intuitive as it is ideal for situations where one does not pay a fine, to be situations where they do not have a student card. This cannot be derived with PT-entailment as we have that the most typical $\neg f$ valuations in $\mathscr{R}_2$ are $\{\neg f, s, \neg p\}$ and $\{\neg f, \neg s, \neg p\}$. The latter valuation means we cannot derive $\bullet\neg f \to \bullet s$.

9.3.3 Exception example

The following is a slight variation of the contrary-to-duty form with the inclusion on an exceptional obligation. Let's say we have the knowledge base $\{\bullet\top \to \bullet\neg p, \bullet p \leftrightarrow \bullet d, \bullet p \to \bullet f\}$. We read these obligations as *"You should not park on campus"*, *"If you have a parking disc then you should park on campus"* and *"If you park on campus then you should pay a fine"*. The PT-entailment models are now given with the ∞-ranked valuations omitted and $\mathscr{R}_1$ being the LM-entailment model.

$\mathscr{R}_1$	0	$\{\neg p, \neg d, \neg f\}, \{\neg p, \neg d, f\}$

$\mathscr{R}_2$	2.	$\{p, \neg d, \neg f\}, \{\neg p, d, f\}, \{\neg p, d, \neg f\}, \{p, d, \neg f\}, \{p, \neg d, f\}$
	1.	$\{p, d, f\}, \{\neg p, \neg d, f\}$
	0.	$\{\neg p, \neg d, \neg f\}$

The LM-entailment model, $\mathscr{R}_1$, gives us $\neg p$ and $\neg d$ as tautologies, which are facts that are read as *"You did not park on campus"* and *"You do not have a parking disk"*. And we see, once again, that PT-entailment blocks these facts since $\mathscr{R}_2$ has valuations which have both p and d being true. A similar explosion pattern to the previous examples is present even in this exception scenario, as taking $\bullet p$ as the antecedent allows any consequent to follow since there is no p-satisfying valuation in $\mathscr{R}_1$. The single additional PT-entailment model, $\mathscr{R}_2$, has valuations where p holds and thus blocks this explosion pattern. While $\mathscr{R}_2$ blocks these above derivations, it still allows for the desirable derivations, by PTL-entailment, of $\bullet p \rightarrow \bullet f$ and $\bullet p \rightarrow \bullet d$(one side of the exception's double implication with p as the antecedent) from the knowledge base.

9.4 Conclusion

We reused the representation from the LM-entailment analysis in this chapter as it uses the same logic language of PTL. We also retained the applicability of the properties from the LM-entailment analysis, since the LM-entailment models used in the properties part of that chapter also qualified as PT-entailment models. When analysing the effectiveness of the PTL-entailment method, we have seen that applying the PT-entailment method produces more conservative derivations than LM-entailment. We were unable to perform analysis on the paradoxes using PT-entailment since the structure of the paradoxes prevented us to do so. This resulted in the construction of new examples which were meant to convey much of the intuition of the paradoxes' contrary-to-duty readings. We built these examples in such a way as to allow the introduction of the additional models required to use the PT-entailment reasoning method. PT-entailment then blocked some unreasonable derivations that LM-entailment produces for these new contrary-to-duty scenarios. Using the representation from the LM-entailment chapter also meant a representation for exceptions was adopted along with the method to detect violations.

Chapter 10

Conclusions And Future Work

This final chapter will firstly summarise the research conducted throughout the dissertation and concisely outline the results. There will then be a brief discussion on potential work that can be performed that follows on from the work done in this dissertation.

10.1 Conclusions

The main goal of this dissertation, as stated in chapter 1, was to analyse how effective defeasibility and typicality could be within a deontic environment. In chapter 2, we presented propositional logic which formed the foundation of the various logic systems that we examined in the dissertation. Chapter 3 dealt with KLM-style defeasible reasoning, which was the chosen logic to start the deontic analysis. We defined the language of KLM-style defeasible reasoning and afterwards, the definition of its semantics, which are based on ranked interpretations, was given. Subsequently, we listed the properties which define a rational entailment relation. The end of this chapter saw the detailing of two reasoning methods which could be used during the analysis, rational closure and lexicographic closure. Deontic logic was formally presented in chapter 4. This started with defining the language of dyadic standard deontic logic (DSDL) which was the specific deontic logic system of our choice. The outlining of the DSDL semantics, based on minimal models, followed and the similarities to the KLM-style ranked interpretations were drawn. The chapter then saw the presentation of the deontic properties, as seen throughout the literature, which were deemed to be desirable in a deontic setting. This also brought us to the outlining of deontic paradoxes which frequently occur in the literature. These paradoxes along with the deontic properties were what we utilised during the analysis. In chapter 5, we presented propositional typicality logic (PTL) which was the second logic system to be examined in the analysis. We defined the PTL language as well as its semantics which also closely resemble that of KLM-style defeasible reasoning and DSDL. The two reasoning methods of PTL were then given, LM-entailment and PT-entailment.

Chapter 6 gave the steps taken during the analysis of each logic. This chapter touched on each of the three analysis stages which dealt with representation, deontic property satisfiability/applicability and deontic paradox resolution, respectively. The analysis began in chapter 7 with the KLM approach to defeasible reasoning. How we represented deontic statements using the KLM approach's defeasible implication operator was stated first. We then showed the deontic properties that are satisfied by the KLM approach while presenting an alternative to the Restricted Strengthening of the Antecedent property. The selection of the lexicographic closure algorithm to conduct the paradox analysis in favour of the rational closure algorithm was then motivated. We also detailed the method to detect violations, which made use of the consistency check of the lexicographic closure algorithm. What followed was the analysis of the paradoxes which showed that the KLM approach was indeed useful in solving the issues within the paradoxes. From the analysis of the Fence paradox specifically, a representation for exceptions was still missing for the KLM approach. Chapter 8 followed a similar pattern to that of chapter 7 except this chapter now dealt with the LM-entailment reasoning method of PTL. We started with a guide with which to translate deontic statements into PTL using the PTL bullet operator. Much like with the KLM approach, our representation method had a similar solution for violation detection by checking for the existence of knowledge-base-satisfying valuations. The deontic properties were then examined within PTL using our chosen representation method and the LM-entailment method. We showed, using various LM-entailment models, that the properties are applicable within PTL. What followed was the analysis of the paradoxes which showed that LM-entailment method was also effective in solving the paradoxes' issues. Chapter 9 was the last analysis chapter, and it examined the PT-entailment method of PTL. The method to translate deontic statements into PTL was carried over from the LM-entailment analysis chapter. The section of the applicability of the properties was also kept brief as the models used for this analysis would have been identical to that of chapter 8. The violation detection method from chapter 8 was also retained. Since PT-entailment could not provide additional models for the paradoxes beyond their single LM-entailment model, we did not check the paradoxes using PT-entailment but rather, new examples were constructed in order to perform the PT-entailment analysis. The basis of this analysis is an attempt to observe the differences in derivations between LM-entailment and PT-entailment. The results showed that PT-entailment provides better results for the newly made examples. Overall, it would seem that defeasibility and the stronger notion of typicality can be effective when used in deontic scenarios, at least when KLM-style defeasible reasoning and PTL are used. It is important to note that there are some drawbacks to these methods which mean these cannot immediately be taken to be more effective than deontic logic. PTL, for instance, is still only applicable in specific scenarios and readings such as the one for translating exceptions, is not the most intuitive. There is a need for further study, such as assessing more complicated and multi-layered examples, in order to definitively state where these notions stand regarding deontic logic. However, the success, when applied to the problematic deontic paradoxes, does bode well for the applicability of these notions when applied to more normative reasoning tasks.

10.2 Future Work

There are a few explorable avenues which follow on from different parts of this research. The first being a method of representing exceptions when using the language of the KLM approach. A reasonable representation was identified using the PTL language but one was not so obvious using the KLM approach. Potentially, one could explore the development of a proof system which shows which of the deontic properties are generally satisfied within PTL given our chosen representation. This would go a step further than this research in which we show that the properties can be applied in particular scenarios with propositional letters. Given the reasonable success of using typicality in the deontic setting, it might be a valuable exercise to examine how typicality could be lifted onto a more expressive logic with the family of description logics providing potential candidates. With a more expressive language, the benefits of typicality will be more applicable to a greater number of scenarios and not just those that fall within the limit of our chosen logic systems' expressive power. It may also be of value to study the potential use of the stronger reading provided by PTL but in a defeasible sense. Concerning examples such Chisholm's paradox and the Fence paradox, the study of multi-level violations and exceptions on this paradoxes respectively, may present a more significant challenge to the methods presented in this dissertation. The development of software tools which apply some methods within this dissertation is also a potential path to go down. Such software could allow for individuals to examine legal reasoning scenarios with much greater ease than if they had to conduct the reasoning manually. This would be especially useful when one has an extensive set of normative reasoning statements that they wish to analyse.

Bibliography

[1] Booth, R., Casini, G., Meyer, T., and Varzinczak, I. On the entailment problem for a logic of typicality. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (2015). 3, 7, 12, 18, 21, 35, 36, 37, 39, 40

[2] Casini, G., Meyer, T., Moodley, K., and Varzinczak, I. Towards practical defeasible reasoning for description logics. In *Proceedings of the 26th International Workshop on Description Logics* (2013), pp. 587–599. 11

[3] Casini, G., and Straccia, U. Rational closure for defeasible description logics. In *JELIA 2010* (2010), vol. 16 of *European Conference on Logics in Artificial Intelligence*, pp. 77–90. 2, 11

[4] Casini, G., and Straccia, U. Lexicographic closure for defeasible description logics. In *Proc. of Australasian Ontology Workshop* (2012), vol. 969, pp. 28–39. 2, 3, 17, 19, 21

[5] Danielsson, S. *Preference and Obligation.* Uppsala, Filosofiska Föreningen, 1969. 76

[6] Gallier, J. H. *Logic for computer science: foundations of automatic theorem proving.* Courier Dover Publications, 2015. 6, 8

[7] Giordano, L., Gliozzi, V., Olivetti, N., and Pozzato, G. L. Semantic characterization of rational closure: From propositional logic to description logics. *Artificial Intelligence 226* (2015), 1–33. 3, 15, 16

[8] Goble, L. *Prima Facie Norms, Normative Conflicts, and Dilemnas.* College Publications, October 2013, ch. 4, pp. 241–352. 24, 27, 28

[9] Grossi, D., and Rotolo, A. Logic in the law: A concise overview. https://pdfs.semanticscholar.org/e1d1/5fe4cf3c29cb5417b6c89368f6c790fe3c60.pdf, 2011. Accessed: May 2019. 1, 3

[10] Hansson, B. An analysis of some deontic logics. *Noûs 3*, 4 (1969), 373–398. 24

[11] Hilpinen, R., and McNamara, P. *Deontic Logic: A historical Survey and Introduction*. College Publications, October 2013, ch. 1, pp. 3–136. 22, 30, 31

[12] Klement, K. Propositional logic. https://www.iep.utm.edu/prop-log/#SH3a. Accessed: September 2018. 1, 6

[13] Kraus, S., Lehmann, D., and Magidor, M. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence 44* (July 1990), 167–207. 3, 10, 12, 49

[14] Lehmann, D. Another perspective on default reasoning. In *Annals of Mathematics and Artificial Intelligence* (March 1995), vol. 15, pp. 61–82. 18, 21

[15] Lehmann, D., and Magidor, M. What does a conditional knowledge base entail? In *Artificial Intelligence* (1992), vol. 55, pp. 1–60. 11, 12, 15, 16

[16] Luenendonk, M. Legal reasoning. https://www.cleverism.com/lexicon/legal-reasoning-definition/, August 2016. Accessed: September 2018. 1

[17] Makinson, D., and van der Torre, L. Input/output logics. *Journal of Philosophical Logic 29* (2000), 383–408. 22

[18] Matek, W., and Truszczynski, M. Autoepistemic logic. *Journal of the Assoclatlon for Computing Machinery 38* (July 1991), 588–619. 10

[19] McCarthy, J. *Circumscription - A form of non-monotonic reasoning.* PhD thesis, Stanford Artificial Intelligence Laboratory, February 1980. 10

[20] McNamara, P. Deontic logic. https://plato.stanford.edu/archives/sum2019/entries/logic-deontic/, 2019. Accessed: October 2019. 28, 47

[21] Parent, X., and van der Torre, L. Detachment in normative systems: Examples, inference patterns, properties. *IFCoLog Journal of Logic and its Applications 4*, 9 (2017). 1, 22, 24, 27, 28, 29, 32, 43, 74

[22] Parent, X., and Van Der Torre, L. *Introduction to Deontic Logic and Normative Systems*. College Publication, UK, 2018. 8, 22, 23, 25

[23] Pigozzi, G., and van der Torre, L. Multiagent deontic logic and its challenges from a normative systems perspective. *IFCoLog Journal of Logic and its Applications 4*, 9 (2017), 2929–2994. 1, 22, 23, 24, 27, 28, 29, 32, 43, 52, 74

[24] Reiter, R. A logic for default reasoning. *Artificial intelligence 13*, 1-2 (1980), 81–132. 10

[25] Russell, S. J., and Norvig, P. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016. 6, 8

[26] Rönnedal, D. Contrary-to-duty paradoxes and counterfactual deontic logic. In *Philosophia* (January 2019). 28

[27] van der Torre, L. *Reasoning About Obligations, Defeasibility In Preference-based Deontic Logic*. PhD thesis, Erasmus University Rotterdam, 1997. 2, 24, 28, 31, 32, 33, 34, 43