A Study of Defeasible Logics

by

Frederick W. Maier

(Under the direction of Robert Robinson)

Abstract

Nute's Defeasible Logic is analyzed relative to semantics for logic programs. Fixpoint semantics are created for two prioritized defeasible logics: one for the ambiguity blocking logic NDL and one for the ambiguity propagating logic ADL. The proof systems for NDL and ADL are shown to be sound with respect to their counterpart semantics. For finite and locally finite theories, they are complete. Under a natural translation of defeasible theory into logic program, the semantics of ADL is shown to correspond under certain restrictions to the Well-Founded Semantics for logic programs. A further technique of translating logic programs into defeasible theories is given; the translation is linear, modular, and preserves the well-founded model of the original program. Antimonotone operators for ADL and NDL are defined which allow the bottom-up calculation of consequences under ADL and NDL. The operators can be used to define an Answer-Set Semantics for defeasible logic. It is also shown that for theories with transitive priorities, ADL and NDL satisfy versions of Cut and Cautious Monotony. Neither logic allows reinstatement. Two further ambiguity propagating logics with reinstatement are defined. One of the logics is unprioritized but is closer to the WFS, while the other incorporates priorities.

Index words: Nonmonotonic Reasoning, Defeasible Logic, Logic Programming, Well-Founded Semantics

A Study of Defeasible Logics

by

Frederick W. Maier

B.A., Spring Hill College, 1996

M.A., Tulane University, 1999

M.S., University of Georgia, 2002

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2007

A STUDY OF DEFEASIBLE LOGICS

by

FREDERICK W. MAIER

Approved:

Major Professor: Robert Robinson

Committee: Donald Nute
Walter D. Potter
Charles Cross

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2007

DEDICATION

To the many kind people I have met in Athens. “These years were real and defining.”

ACKNOWLEDGMENTS

I express sincere gratitude to Drs. Nute and Robinson for serving as my (unofficial and official) advisors. I surely would not have been able to complete the degree without their encouragement, guidance, and general good will. Similar thanks go to Drs. Don Potter, Charles Cross, and Budak Arpinar for bearing witness to much of the research. Also, since serving on a dissertation committee is at times like jury duty, they all have my sympathies as well.

I also thank UGA's Department of Computer Science and especially the Artificial Intelligence Center for their longstanding support. UGA is perhaps the most stimulating and amicable working environment I have encountered, and although I have attended several schools over the years, I feel the most loyalty to it.

Special thanks go to David Billington, who read a draft of this dissertation and provided detailed comments. The dissertation is much improved as a result.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This dissertation is a study of defeasible logic, a family of nonmonotonic logics originally developed by Donald Nute in the mid-1980's [Nut86][Nut87]. The logics share a simple rule-based language and are motivated by the intuition that some inferences, such as that involved in

1. Joe is a dog.
2. Richard normally likes dogs.
3. Therefore, Richard likes Joe.

are *defeasible*. In particular, (2) above can be considered a rule of thumb that might be defeated by other evidence. If it is also known that Richard universally dislikes dogs that eat his shoes and that Joe has done just that, then (3) is not a valid conclusion in any of Nute's defeasible logics.

A consequence relation $\vdash$ over propositions is said to be *monotonic* just in case the following holds.

$$\text{If } A \vdash p, \text{ then } A \cup B \vdash p.$$

A and B are sets of propositions and p is a proposition. The consequence relations associated with defeasible logic do not satisfy this condition, and so the logics are called *nonmonotonic*. The symbol $\mid\!\sim$, first used in [Gab85], is generally used to indicate nonmonotonic consequence. One may read $A \mid\!\sim p$ as stating that p is defeasibly derivable from premises A. Defeasible logic defines a counterpart relation $\sim\!\mid$ which may be interpreted as "defeasible refutation". $A \sim\!\mid p$ states that p is *demonstrably* not derivable from A. That is, all arguments supporting p have been shown to

fail. Note that $A \sim\!\!\mid p$ does not mean $A \not\sim p$, though in any reasonable logic $A \sim\!\!\mid p$ should imply $A \not\sim p$. A few words of Donald Rumsfeld are (surprisingly) helpful.

> There are things we know that we know. There are known unknowns. That is to say there are things that we now know we don't know. But there are also unknown unknowns. There are things we don't know we don't know. So when we do the best we can and we pull all this information together, and we then say well that's basically what we see as the situation, that is really only the known knowns and the known unknowns. And each year, we discover a few more of those unknown unknowns [Rum02].

Defeasible logic similarly partitions propositions into three sets. There are "known knowns" ($D \mid\!\sim p$). There are "known unknowns" ($D \sim\!\!\mid p$). But there are also "unknown unknowns" ($D \not\sim p$ and $D \not\sim\!\!\mid p$). Furthermore, the discovery of new information necessarily causes the lines delineating the three to shift.

The research for this dissertation had two general goals in mind: (1) to investigate the formal properties of Nute's defeasible logic, and (2) to determine its relationship to other nonmonotonic reasoning (NMR) formalisms, particularly logic programming under the well-founded semantics (WFS) [GRS91]. The motivation for the research was a belief that, compared to other NMR formalisms such as default logic or logic programming, defeasible logic was less understood. The research was intended to better pinpoint defeasible logic's position on the NMR map. The well-founded semantics was chosen because it is, like defeasible logic, a directly skeptical formalism (direct skepticism is defined in section 1.2) and bears at least a superficial resemblance to Nute's most recent logic [Nut01]. Furthermore, the logic of [Nut01] had never before been examined relative to logic programming.

The major results of this dissertation are summarized at the very end of this chapter. Before we state them, however, we will discuss nonmonotonic reasoning in a rather general way. This is done in order to make the results and the remainder of the dissertation more understandable.

1.1 NONMONOTONIC REASONING

Nonmonotonic reasoning is often characterized by the phrase "jumping to conclusions": it is the drawing of conclusions that are not deductive consequences of knowledge currently possessed. This characterization is true as far as it goes, but it fails to mention the essential facts about NMR that distinguish it from patterns of reasoning that are merely bad. Specifically it fails to mention that (1) the conclusions drawn are nevertheless justified by the information currently possessed, and (2) they might later be retracted once additional information is obtained (in other words, nonmonotonic reasoning is nonmonotonic). It is clear that certain patterns of human reasoning are in fact nonmonotonic, and they can't be considered bad even if they are not exactly classical. In truth very little in real life can be justified using classical logic. Quine's phrase, "the Humean predicament is the human predicament", applies. Nonmonotonic logics attempt to capture these patterns of reasoning in a symbolic formalism.

Historically, NMR formalisms have been tied to a traditional school of artificial intelligence that proposes human-like common sense can be achieved in a machine provided a suitable logical formalism is found. Certainly, much of the early discussion of NMR was done in the context of this strong AI. According to Minker [Min93], one of the earliest publications to stress the need for implementing common sense reasoning for artificial intelligence is a 1959 paper by McCarthy, [McC59], which McCarthy himself admits to likely be the first paper on logic-based AI.[1] Hayes (1973) states the need for nonmonotonicity in a discussion of the frame problem [Hay73]. Minsky in [Min74] is apparently the first to actually use the term 'monotonicity' in conjunction with logic-based AI systems; in particular, he states that the monotonic nature of classical logic would be a significant impediment to its use in AI. Implementations of nonmonotonic logical systems appeared in the early 1970's, though the developers might not have been entirely aware that this is what their systems were. Prolog, which interprets "$not\ \ p$" as failure to prove p, appeared in 1973 and is clearly nonmonotonic (for an interesting history of Prolog's inception, see [CR93]). Reiter's formulation of the closed world assumption (CWA) for databases appeared in 1977–1978 [Rei77].

[1] See http://www-formal.stanford.edu/jmc/mcc59.html. Accessed May 7, 2007.

McCarthy's Circumscription and Reiter's Default Logic were formulated in some form at roughly this time. Articles on both appeared in a 1980 volume of *Artificial Intelligence* [Rei80] [McC80], as does an article by McDermott and Doyle introducing a nonmonotonic modal logic [MD80]; the dominant formalism based on the same idea is Moore's autoepistemic logic [Moo85].

One needn't espouse the view that strong AI is possible, however, to feel that a nonclassical logic might be useful or interesting. Computers are labor saving devices, and at least some of the labor automated in them is the mental labor of drawing inferences and making decisions. Furthermore, machines are forced to work in as information–deprived an environment as humans do and so must reason nonmonotonically at least some of the time. It's usually the programmer, however, who is explicitly instructing the machine what to do in the absence of information or in the presence of inconsistencies. Often, this is done in an ad hoc manner. NMR systems can be viewed as an attempt to make this process slightly more systematic.

1.2 Basic Concepts in NMR

Though there are many nonmonotonic formalisms, several concepts occur repeatedly in the NMR field as a whole. Since they bear a direct relation to the discussion of defeasible logic in this dissertation, we present them here. They are most easily understood via illustrative examples such as the one shown below.

Example 1.1. *The Nixon Diamond.*

1. *Nixon is a republican.*
2. *Nixon is a Quaker.*
3. *Quakers are normally pacifists.*
4. *Republicans are normally not pacifists.*

We shall call the whole set of statements above a *theory*. The first two statements are taken to be strict facts of the world. The latter two statements are defeasible and in conflict. A person cannot be both a pacifist and a non-pacifist. The example might be encoded in Nute's defeasible logic as

1. $\{\} \rightarrow R$
2. $\{\} \rightarrow Q$
3. $\{Q\} \Rightarrow P$
4. $\{R\} \Rightarrow \neg P$

All statements are represented as rules. Defeasible logic uses '$\rightarrow$' to denote strict statements and '$\Rightarrow$' to denote defeasible ones. The indisputable facts that Nixon is a Quaker and a Republican are encoded as rules with empty antecedents. The distinction between strict and defeasible information can be found elsewhere. A default theory of Reiter's default logic [Rei80] is a tuple $\langle W, D \rangle$, where W is a set of first-order formulas representing hard facts of the world and D is a set of *defaults* representing the defeasible information. Strict and defeasible rules can also be found in the work of Pollock [Pol87], Prakken and Sator [PS97], and in certain forms of inheritance networks [Hor94]. The typical representation of the example as an inheritance network is shown in Figure 1.1, where nodes indicate classes and edges indicate strict or defeasible membership in a class. The edge $Q \Rightarrow P$ indicates that, defeasibly, a member of Q is a member of P. The edge $R \not\Rightarrow P$ means that, defeasibly, a member of R is not a member of P. It is the analog of the defeasible rule $R \Rightarrow \neg P$. The node N may be taken as the designated starting point. In the example, all members of N are strictly members of Q and R.

In the network, the path $N \rightarrow Q \Rightarrow P$ constitutes an *argument* with conclusion P, while $N \rightarrow R \not\Rightarrow P$ is an argument with conclusion $\neg P$ (analogous paths can be constructed in the context of defeasible logic). The arguments, since they have complementary literals as conclusions, are in conflict; they are sometimes said to *attack* each other. Or, in the language of Pollock [Pol87], each is a *rebutting defeater* of the other.

Given a theory such as the Nixon diamond, the basic issue is to determine the conclusions that can be justifiably drawn. One might conclude that Nixon is a pacifist because he is a Quaker. This amounts to taking the left path in the diagram. Alternatively, because Nixon is a Republican, one might conclude that Nixon is not a pacifist (taking the right path). Or, not seeing a good reason to choose one over the other, one might refuse to take either path. The various nonmonotonic

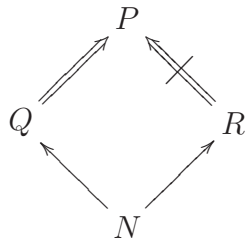

Figure 1.1: A Typical Example: The Nixon Diamond

formalisms are intended to delineate the set of justified consequences, and they often disagree over what this set actually contains.

Many of the formalisms—including default logic, autoepistemic logic, inheritance networks, the stable model/answer set semantics for logic programs [GL88][GL91], and Dung's argumentation framework [Dun95]— define the consequences in terms of *extensions*, though each formalism might define extension differently or give it another name. Regardless of the formalism, however, the idea is basically the same. Informally, an extension is a rational and stable set of beliefs that one might espouse based upon the theory. An extension is rational in the sense that defeasible conclusions are not accepted if they introduce inconsistencies. In the example, one should not accept both P and $\neg P$. An extension is stable in the sense that (1) everything in the extension is 'justified' by the theory, and (2) nothing more can be added without introducing some further inconsistency. In the Nixon example, the sets (1) $\{N, Q, R, P\}$, and (2) $\{N, Q, R, \neg P\}$ each constitute an extension of the theory. Each is a consistent set and nothing more can be added without introducing inconsistency. The set $\{N, Q\}$ is not a valid extension because it can be consistently extended with R. The set $\{N, Q, R, P, \neg P\}$ is not an extension because it is inconsistent. $\{N, Q, R, \neg P, E\}$ is not an extension because E is not supported at all by the theory. Depending upon how extensions are defined in a particular nonmonotonic formalism, some theories might have many extensions, or a unique one, or perhaps none at all. ***E.g.***, in the context of inheritance networks without strict

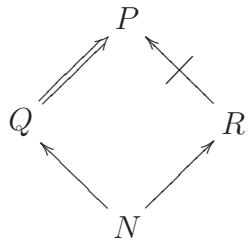

Figure 1.2: Nixon is certainly not a pacifist.

links, extensions are simply maximal conflict free subgraphs (rooted at some designated node such as N), and so every network has at least one extension.

If the diamond from the example is changed (as in Figure 1.2) so that the link between Republicanism and anti-pacifism is strict, then in defeasible logic, as in most formalisms, there is only one valid extension. In it Nixon is a Quaker, a Republican, and not a pacifist. The argument $N \rightarrow R \not\rightarrow P$ is said to *defeat* the argument $N \rightarrow R \Rightarrow P$. The same effect can be achieved if one places priorities on defeasible arguments, so that one argument is preferred to another. Either way, the notions of conflict and defeat are thus potentially asymmetrical.

Returning to the original Nixon diamond and the conclusions that should be drawn from it, the distinction may be made between *credulous* (or *brave*) reasoning and *skeptical* (or *cautious*) reasoning. With credulous reasoning, a proposition p is a consequence of the theory if it is contained in *any* extension (or perhaps some *preferred* extension). One of the issues in accounts of credulous reasoning is how to justify, if one extension is to be preferred to another, *why* it is preferred to another. In the answer set semantics for logic programs, no preference exists. Extensions, called in that context *answer-sets*, are intended to model possible solutions to a problem or provide alternative diagnoses. Strictly speaking, no answer set is preferred to another.

The alternative is skeptical reasoning. In most formalisms, the skeptical consequences of a theory are taken to be the statements that are members of every extension. In the present example, the set $\{N, Q, R\}$ constitutes the skeptical consequences of the theory. The literal P, because conflicting arguments exists for it (alternatively, because it exists in some extensions but not others),

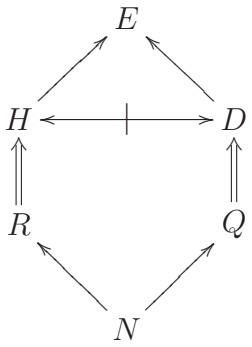

Figure 1.3: Floating Conclusions: Doves and Hawks

is said to be *ambiguous* [Ste89]. Stein has called ambiguity an obligatory part of nonmonotonic reasoning and its proper handling one of the central issues [Ste90]. This should be clear, since ambiguity is the direct result of conflict, and conflict resolution *is* central to nonmonotonic reasoning.

Defining skeptical consequences as the intersection of sets of literals is not the only available option, however. Alternatively, one might build acceptable arguments incrementally and take the conclusions of acceptable arguments as the consequences of the theory. Informally, one might consider an argument acceptable if there is no equally "good" argument that conflicts with it. All forms of defeasible logic build sets of consequences in this incremental fashion.

The two forms of skepticism are not identical. Consider the below example, symbolized as an inheritance net in Figure 1.3.

Example 1.2.

1. *Nixon is a republican.*
2. *Nixon is a Quaker.*
3. *Quakers are normally Doves.*
4. *Republicans are normally Hawks.*
5. *Hawks are never Doves, and vice versa.*
6. *Hawks are Political Extremists*
7. *Doves are Political Extremists*

There are two extensions to this theory: $\{N, Q, R, D, \neg H, E\}$ and $\{N, Q, R, H, \neg D, E\}$. The intersection of these two sets is $\{N, Q, R, E\}$, and so under the first notion of skeptical consequence, Nixon is an extremist regardless of whether he is a hawk or a dove. However, under the second form of skepticism, the argument $N \rightarrow Q \Rightarrow D$ would not be acceptable since there is a competing argument with conclusion $\neg D$ and no way to defuse the conflict. Because of this, expanding the argument to $N \rightarrow Q \Rightarrow D \Rightarrow E$ is not allowed. A similar state of affairs obtains for the argument $N \rightarrow R \Rightarrow H \Rightarrow E$, and so E is not a conclusion of the theory.

Formalisms that espouse the first view are called *indirectly skeptical* because the consequences are determined only after all extensions have been generated. They are also sometimes called *ideally skeptical*. Formalisms based on the latter view are called *directly skeptical* because the acceptable arguments and their conclusions are built directly and without first taking the intersection of extensions. The justification for the directly skeptical view is that the reasons for believing a proposition (say, that Nixon is an extremist) should themselves be justified.

In the example, "Nixon is politically extreme" is called a *floating conclusion*. Perhaps the majority of researchers view indirect skepticism as more reasonable than direct skepticism, and perhaps because of this they accept floating conclusions as perfectly reasonable. Stein writes, "The failure of these path-based approaches to be both sound and complete for ideally skeptical inheritance indicates the importance of reasoning about conclusions, or inferences, rather than about their supporting paths, or arguments" [Ste89]. Makinson and Schlecta call the directly skeptical approach "just wrong" [MS91]. In their view, direct skepticism is *at best* a tractable approximation to the much more rational and computationally expensive indirect skepticism This, it is noted, is precisely what is said of the well-founded semantics [GRS88] for logic programs in comparison to the stable model semantics [GL88]. The directly skeptical well-founded semantics approximates the indirectly skeptical stable-model semantics.

Acceptance of floating conclusions is not universal. All versions of defeasible logic are directly skeptical and reject them. Horty takes issue with them as well; an informal argument against them is shown below.

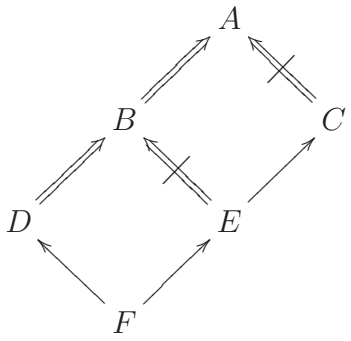

Figure 1.4: The Double Diamond.

> Of course, Nixon's own position on the matter is well known. But if I were told of some other individual that he is both a Quaker and a Republican, I would not be sure what to conclude. It is possible that this individual would adopt an extreme position, as either a dove or a hawk. But it seems equally reasonable to imagine that such an individual, rather than being pulled to one extreme or the other, would combine elements of both views in a more balanced, measured position falling toward the center of the political spectrum[Hor02].

1.2.1 Ambiguity and Zombie Paths

Consider Figure 1.4 (found in [THT87] [MS91]). There are three extensions: $\{F, E, D, C, B, A\}$, $\{F, E, D, C, B, \neg A\}$, and $\{F, E, D, C, \neg B, \neg A\}$. The intersection is simply $\{F, E, D, C\}$. The Literal B is ambiguous in the sense described in the previous section. Many researchers (the majority, it seems), perhaps because they already espouse indirect skepticism, view that the ambiguity of B should be *propagated* to all statements that depend upon it. A formalism which implements this view is called *ambiguity propagating*. In the example, A would be ambiguous as well. Propagating ambiguity potentially affects the conclusions drawn from the theory. Makinson and Schlecta call the path to $F \rightarrow D \Rightarrow B \Rightarrow A$ a *zombie path* [MS91]. The path is "dead" in the

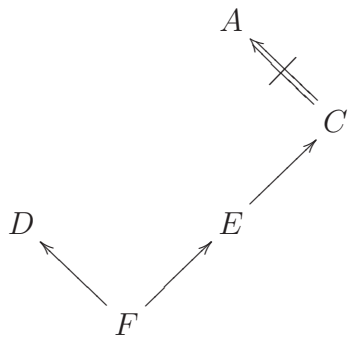

Figure 1.5: Blocking Ambiguity.

sense that one can't accept its conclusions; however it is still "alive" enough to conflict with the path $F \rightarrow E \rightarrow C \not\Rightarrow A$, and this prevents one from concluding $\neg A$.

The opposing view, called *ambiguity blocking*, takes B as refuted. This is the view espoused by Horty in [HTT87] and by most forms of Nute's defeasible logic. The justification is that since there is no clearly acceptable evidence for B, it should be rejected. This is tantamount to deleting from the inheritance network node B and all edges to and from it (this is shown in Figure 1.5). If that is done, then all support for A disappears, leaving only support for $\neg A$. In 1987 Pollock admitted defeasible logic to be the only formalism he could think of that blocks ambiguity (the major papers on inheritance networks (*e.g.*, [HTT87]) appeared at roughly the same time)[Pol87]. The issue of ambiguity is discussed in some detail by Stein in [Ste90] and [Ste89].

Adopting ambiguity propagation yields in a sense a more extreme form of skepticism, in that *fewer* conclusions can be drawn. In the example, B *might* hold (there is conflicting information about it and no way to resolve the conflict), and if it does hold, then there would be evidence for both A and $\neg A$. In the ambiguity propagating view, the safe course of action is to conclude nothing at all about A.

1.2.2 REINSTATEMENT

In the examples considered thus far, we have looked at only very simple forms of conflict. The matter can be more complicated, however. Consider the following story: Joe lives alone, and men who live alone are usually bachelors. Bachelors are by definition not married. But it is known that Joe is in fact married (his wife attends school in a distant state). This may be represented as the below theory in defeasible logic.

Example 1.3.

1. $\{\} \rightarrow livesAlone(joe)$
2. $\{\} \rightarrow man(joe)$
3. $\{man(X), livesAlone(X)\} \Rightarrow bachelor(X)$
4. $\{bachelor(X)\} \rightarrow \neg married(X)$.
5. $\{\} \rightarrow married(joe)$.

Intuitively, the fact that Joe lives alone defeasibly implies that he is a bachelor. But this is contradicted by the fact that he is married. Rules 3 and 5 are in conflict, even though rule 3 doesn't explicitly mention anything about being married; that is, the conflict is indirect. Furthermore, since rule 5 is strict and rule 3 only defeasible, rule 5 should win out over rule 3. In Prakken and Sator's terminology [PS97], 5 *strictly defeats* 3, since 5 defeats but is itself undefeated. The example shows that when hard facts (rules 4 and 5) are used in conjunction with defaults (rule 3), determining and resolving conflict is not always easy.

Suppose that 5 above is not absolutely certain, but that the evidence for Joe being married is very strong—*e.g.*, he has a seemingly valid marriage certificate and we consider this to be stronger evidence than the correlation between bachelorhood and living alone. In this modified case, the correct result is still to conclude that Joe is married. However, if new information comes to light that runs contrary to him being married (for instance, the document is forged), the situation is substantially different. There is now an argument that Joe is a bachelor (he is one because he lives alone), an argument that he is married (there is a marriage certificate), and a further argument that he is not married (the document is fake). Each argument appears stronger than the last.

Many researchers feel that, once rule 5 has been discounted, the argument that Joe is a bachelor should be *reinstated*. Rule 3 should hold, and using it we might be able to draw further conclusions. Horty, Touretzky, and Thomason, however, appear completely opposed to reinstatement [TTH91] [Hor01]. They give several examples which indicate that reinstatement is unreasonable. For instance, while it is true that chickens with jet packs can fly, this has nothing to do with the fact that chickens are birds, and so why should one allow a rule like $jetpack \rightarrow flies$ to reinstate a rule like $bird \Rightarrow flies$ which is defeated by another $chicken \Rightarrow \neg flies$? In this case, the reinstatement doesn't much matter ($flies$ is concluded in either case). But in other cases, a reinstated rule allows conclusions to be drawn that might not otherwise be possible.

Example 1.4. *[Hor01]*

1. r_1*: Microsoft employees are normally millionaires.*
2. r_2*:* Newly *hired Microsoft employees normally have less than $500,000.*
3. r_3*: Beth is a new Microsoft employee and we have very good reason to think she has more than $500,000.*
4. r_4*: millionaires also have at least $500,000.*

In the example r_1 is considered weaker than r_2, and r_2 is considered weaker than r_3. One may simplify the example a bit by assuming that it's a fact that Beth has $500K. Either way, Horty argues that it is incorrect to conclude that Beth is a millionaire. But that's the conclusion that reinstatement implies. r_1 conflicts with r_2, but r_2 is defeated by r_3, and so r_1 is reinstated. Certain defeasible logics (*e.g.*, [Bil93]) allow reinstatement; some ([Nut01][MN06a]) do not.

In the remainder of this dissertation, we will take reinstatement as a reasonable principle to adopt. Few researchers, even if they find Horty's examples troubling, think them sufficient to warrant rejecting reinstatement (see [Pra02] for a defense against Horty). Certainly, in the bachelor example, it seemed reasonable to suppose, once the evidence for his marriage was discounted, that Joe is a bachelor. In general, we feel that an argument that has been defeated should not be used to defeat other arguments. In later chapters, it will be shown that the defeasible logics NDL and ADL do not allow reinstatement; we present alternative logics that do accommodate it.

1.3 Consequence Relations: Formal Properties

In 1989, Horty stated (perhaps jokingly) that there might be 72 different nonmonotonic formalisms and reasoning strategies being investigated [DL91]. Some of these have since died out, but others have sprouted up. Hayes reportedly once said at a workshop that anyone inventing a new logic should be fined $1000.[2] Indeed, there are a great many NMR formalisms. Given this, how are we to evaluate their worth?

In the previous sections, the technique used was to develop test cases such as the Nixon Diamond and examine a given formalism's results against those of another system or perhaps intuition itself. While this technique is commonly used, it is problematic for several reasons, not the least of which is that it's not very systematic. Furthermore, intuitions are very personal in nature; often one person's intuitions clash with those of another. Even more troubling is that a single person's intuitions can vary on two examples that appear formally identical.

An alternative means of evaluation is to consider directly the consequence relations that the formalisms define and determine whether these relations possess certain desirable formal properties. This is the approach taken by Dov Gabbay in a 1985 paper [Gab85] and by others (particularly, Makinson in [Mak94]). Gabbay's paper explicitly addresses what it means for a nonmonotonic *formalism* to constitute a nonmonotonic *logic*. The conclusion he arrives at—one that has been to a large extent accepted—is that any relation $\mid\!\sim$ constitutes a nonmonotonic consequence relation if (and only if) it satisfies the below three properties.

1. **Reflexivity**: If $p \in \Gamma$, then $\Gamma \mid\!\sim p$.
2. **Cut**: If $\Gamma \mid\!\sim p$ and $\Gamma \cup \{p\} \mid\!\sim q$, then $\Gamma \mid\!\sim q$.
3. **Cautious Monotony**: If $\Gamma \mid\!\sim p$ and $\Gamma \mid\!\sim q$, then $\Gamma \cup \{p\} \mid\!\sim q$.

It can be seen that Cut and Cautious Monotony are converses of each other. They are sometimes combined into a single property called Cumulativity.

4. **Cumulativity**: If $\Gamma \mid\!\sim p$, then $\Gamma \mid\!\sim q$ iff $\Gamma \cup \{p\} \mid\!\sim q$.

[2] This story is recounted by Murray Shanahan in [Sha97].

Why are Reflexivity, Cut, and Cautious Monotony viewed as special? If a relation $\mid\!\sim$ satisfies Reflexivity, Cut, and Monotony, then according to Tarski, it defines a consequence relation for some monotonic logic. Gabbay writes: "In fact we can say that to be a deductive monotonic logical system is to be a relation $\vdash$ satisfying the three conditions . . ." [Gab85]. If a similar set of properties is to be used to define nonmonotonic consequence relations, then Monotony must be abandoned. Gabbay recommends it be replaced with Cautious Monotony, which is itself a restricted form of Monotony ('restricted monotony' is precisely what Gabbay calls it).

Even if one rejects the notion that a nonmonotonic consequence relation *must* satisfy Cut and Cautious Monotony in order to count as a logic (Default Logic fails Cautious Monotony [Mak94], as does the stable model/answer set semantics, but most consider these to be nonmonotonic logics), one must see that they are important and desirable properties to satisfy. Cautious Monotony allows lemmatization: if p is a consequence of a given theory, then we may add it to the theory without changing the set of consequences. Cut is its converse: we can remove from the theory the things that are consequences of the remainder without affecting the results. That is, we can safely remove the redundant statements. Together Cut and Cautious Monotony make a logic which is in some sense stable and well-behaved.

Two other significant properties that a nonmonotonic system might possess are shown below.

5. **Supraclassicality**: If $\Gamma \vdash_{FOL} p$, then $\Gamma \mid\!\sim p$.
6. **Consistency Preservation**: If $\Gamma \mid\!\sim p$ and $\Gamma \mid\!\sim \neg p$, then for all q

   $\Gamma \vdash_{FOL} q$ and $\Gamma \vdash_{FOL} \neg q$.

Supraclassicality is readily motivated: it is often desirable that a nonmonotonic reasoning system *extend* classical reasoning systems, allowing one to draw *more* conclusions. Nevertheless, there are good reasons for not wanting it, the most obvious one being that classical reasoning is explosive, in the sense that everything follows from an inconsistency. This means that even irrelevant inconsistencies in a theory prevent one from drawing any useful information at all. Consistency Preservation means that the nonmonotonic machinery of the formalism does not introduce any inconsistencies that aren't already classically entailed. It certainly is a desirable property to have, though it often comes at a computationally high price.

1.4 Contributions of the Dissertation

There have been several distinct defeasible logics developed over the years. Nute's most recent logic, which we call NDL ([Nut01]), bears a resemblance to logic programming under the *well-founded semantics* (WFS) [GRS88] [GRS91]. Much of the research for this dissertation has consisted of studying the relationship between the two. The investigation has born some fruit. Particularly, it allowed for the creation of a semantics for NDL based on the WFS as well as an alternating fixpoint procedure for computing consequences of defeasible theories.

The formal properties of NDL and related logics were also studied. For arbitrary theories, the logics do not satisfy Cut and Cautious Monotony. The properties are satisfied, however, for theories using a transitive priority relation on rules. For arbitrary theories, lesser but related properties are satisfied. It was also discovered that the logics do not allow reinstatement to occur. This observation motivated the development of further logics.

A summary of the dissertation's results appears below.

1. A method of embedding defeasible theories of NDL into logic programs under the well-founded semantics was developed. This has been published in [MN06b]. While not novel (the work is based upon that found in [AM02] and [ABGM06]), the embedding is at least useful. It allows one to perform reasoning under NDL via a logic programming system that supports the WFS.

2. An ambiguity propagating variant (ADL) of NDL was created by modifying the proof system of NDL. ADL is a more conservative logic than NDL and draws fewer conclusions. Where ADL does draw a conclusion, it agrees with NDL. The proof system for ADL has been published in [MN06a].

3. Fixpoint semantics inspired by the WFS have been created for both NDL and ADL. The proof systems are sound *wrt* the new semantics. For finite and "locally finite" theories, they are complete. A more general characterization of the circumstances under which the proof systems and semantics coincide has also been found. The semantics are preferable to the

indirectly skeptical semantics developed by Donnelly [Don99], if only because (1) the proof systems are sound *wrt* the new semantics and for finite theories they are complete (completeness does not hold *wrt* Donnelly's semantics, even for finite theories), and (2) the new semantics are more closely related to the well-founded semantics and also the stable model semantics. The latter point is useful, for it better pinpoints defeasible logic in relation to other NMR formalisms.

4. For arbitrary prioritized defeasible theories, both ADL and NDL fail simple versions of Cautious Monotony and Cut. For theories with transitive priorities, however, these properties do hold. Related properties hold for arbitrary theories; in that sense, ADL and NDL are well–behaved.

5. Under a natural translation of unprioritized defeasible theories into logic programs and under certain other restrictions, the semantics of ADL corresponds to the well-founded semantics for normal logic programs.

6. A modular way of translating logic programs into defeasible theories was also discovered. The well-founded model of the programs correspond to the models of the corresponding defeasible theories under ADL. For this class, the NDL semantics can be viewed as providing an ambiguity blocking semantics for normal logic programs.

7. Two antimonotone operators have been created for defeasible logic. One is ambiguity propagating and to a large extent parallels ADL; the other is ambiguity blocking. For theories lacking priorities, the model generated by the ambiguity propagating operator coincides with the well–founded model of the theory's logic program translation. The model produced by the ambiguity blocking operator coincides with the model under NDL, even for theories with priorities. These operators can also be used to define a stable model/answer-set semantics for defeasible logic. It turns out, in fact, that the ambiguity propagating operator corresponds exactly to the Gelfond-Lifschitz operator for stable models.

8. It was found that NDL and ADL do not allow reinstatement of arguments, and for a large class of defeasible theories this leads to very unintuitive results. Two logics were created which avoid these problems. One does not incorporate priorities but is structurally very close to the WFS. The other does incorporate priorities.

1.5 Outline of Later Chapters

The remaining chapters discuss these results in some detail. Chapter 2 provides an overview of first-order logic, the syntax of logic programming, and presents the two dominant semantics for logic programs: (1) The stable model/answer set semantics and (2) the well-founded semantics. Subsequent chapters use this material to relate defeasible logic to these semantics.

Chapter 3 introduces the defeasible logic paradigm and the proof-systems for the major defeasible logics: (1) BDL (and its offshoots), the variant extensively researched by David Billington and others, and (2) NDL and its ambiguity propagating counterpart ADL. There are several points of difference between the logics in (1) and those in (2). Nevertheless it is shown that to a certain extent features can be swapped between the logics.

Chapter 4 presents the fixpoint semantics for NDL and ADL. Finite grounded components are defined and proofs of soundness and completeness *wrt* finite grounded components are given. The class of locally finite theories, which includes all finite defeasible theories, is also defined. Since well-founded and unfounded locally finite literals are also members of finite grounded components, it may be inferred that the proof systems for NDL and ADL are sound and complete *wrt* the semantics for locally finite theories. Chapter 4 also presents alternating fixpoint procedures for computing the results of ADL and NDL, and it is shown that both NDL and ADL satisfy Cut and Cautious Monotony if the priorities among rules are transitive but that they fail these properties in general.

Chapter 5 shows how defeasible theories can be translated into logic programs so that, for restricted theories, the consequences under ADL correspond to the well-founded models of their translations. The ambiguity propagating alternating fixpoint procedure from chapter 4 is shown to

exactly correspond to the GL-operator used for the stable model semantics and the WFS. Means of eliminating priorities and defeaters from defeasible theories are also given.

Chapter 6 presents two new ambiguity propagating logics, which we call SDL and MDL, respectively. SDL is more closely related to the WFS—*i.e.* the correspondence holds for a larger class of defeasible theory—and allows reinstatement. However, it does not incorporate priorities (which admittedly is a significant drawback). MDL incorporate priorities and allows reinstatement.

Several appendices are also included. The first presents examples illustrating where various defeasible logics differ. The second gives a means of translating defeasible theories of NDL into logic programs. It has the benefit of incorporating priorities among rules. Later appendices provide proofs for theorems stated but not proven elsewhere in the dissertation.

CHAPTER 2

THE SEMANTICS OF LOGIC PROGRAMS

A logic program consists of a set of "if . . . then" rules of the following form.

$$p \texttt{:-} a_1, a_2, \ldots, a_n,\ not\ b_1,\ not\ b_2, \ldots,\ not\ b_m$$

where the consequent of the rule p and all of the a's and b's in the antecedent of the rule are atomic formulas or their logical complements. The various semantics defined for logic programs specify different consequence relations, and they are all nonmonotonic. Excluding relational database systems, logic programming is the primary nonmonotonic reasoning formalism in computer science today. The non-monotonicity can be seen by considering the program below.

Example 2.1.

1. $p \ :- \ not\ q.$

Informally, the above rule is interpreted simply as "If *not* q, then p." Formally, there are several competing proposals for what the rule actually means. None of the proposals take the '*not*' to be the same as classical negation. If one interprets the rule as the material conditional $\neg q \rightarrow p$, then p is not a classical consequence of the program. However, it is a consequence under virtually every logic programming semantics. The stable model semantics [GL88] considers not to mean "not believed"; in the present program $not\ q$ holds because there is no rule supporting q and so no reason to believe q. Prolog takes not to mean 'not derivable via SLD-Resolution'. In both accounts, p is a consequence of the program. Adding q as a fact changes things, however.

Example 2.2.

1. p :– $not\ q$.
2. q.

Here p is no longer a consequence of the program. The previous conclusion has been withdrawn.

In this chapter we present logic programming concepts relevant to the later discussion of defeasible logic. We pay special attention to the stable model/answer set semantics [GL88][GL91] and the well-founded semantics [GRS91], which are widely viewed as the main semantics for logic programs. However, since logic programming is itself based in large part on classical logic, we first quickly review concepts of first order logic (FOL) and model theory. The below presentation is adapted mostly from the formal system defined in Enderton [End72]. However, any standard presentation would do. We have modified the system to treat $\wedge$ and $\exists$ as primitives and have eliminated the equality symbol $\approx$. We also write terms and atomic formulas in prefix notation using parentheses and commas (for instance, $P(x, y, z)$), as is commonly done in logic programming.

2.1 First Order Logic: Syntax and Semantics

A *formal language* L is a set of strings over some base set of primitive symbols called a *vocabulary* or *alphabet*. In computer science, the alphabet is generally required to be finite and L might not be decidable. However, in the case of logical languages, the alphabet is allowed to be countably infinite and is usually divided into disjoint sets (to distinguish variables, constants, logical connectives, etc.), and L must be a decidable set. The strings $w \in L$ are called *well-formed formulas* (*wffs*) of L.

Definition 2.3. *A* formal system *consists of*

1. *a formal language,*
2. *a set of distinguished wffs to be taken as* axioms*, and*
3. *a* consequence relation *associating sets $S \subseteq L$ to individual members $x \in L$.*

The set of axioms may be empty, but it also may be infinite; if infinite, then it is usually represented by a small set of axiom-schemas (each axiom matches one of the schemas).

Definition 2.4. *The vocabulary of a first order language consists of the following (disjoint) sets:*

1. predicate symbols*:* P_0*,* P_1*, . . ., each with a finite* $arity\ n \geq 1$.
2. constant symbols*:* a_0*,* a_1*, . . . (allowed to be empty).*
3. variables*:* v_0*,* v_1*, . . . (one for each integer* $n \geq 0$*).*
4. logical connectives*:* $\{\wedge, \neg, \rightarrow\}$.
5. quantifier symbols*:* $\{\forall, \exists\}$.
6. punctuation symbols*:* {'(',')'}.
7. function symbols*:* f_0*,* f_1*, . . ., each with a finite* $arity\ n > 0$ *. This set is allowed to be empty.*

The set of constants is allowed to be empty, as is the set of function symbols, but the set of predicates is assumed to have at least one member. In the above, the subscripts are used for informative purposes only. The other common logical symbols $\{\vee, \leftrightarrow\}$ are not taken as primitives. In the following, we will use bold font symbols as meta-variables to talk about predicate symbols constants, function symbols, etc.

Definition 2.5 (Terms)**.** *Terms are defined inductively.*

1. *A variable or constant* $\mathbf{t}$ *is a term.*
2. *If* $\mathbf{f}$ *is an n-ary function symbol and* $\mathbf{t_0}$*,* $\mathbf{t_1}$*, . . .,* $\mathbf{t_n}$ *are terms, then* $\mathbf{f}(\mathbf{t_0}, \mathbf{t_1}, \ldots, \mathbf{t_n})$ *is a term.*

Nothing not fitting one of the above conditions is a term. If a term contains no variables, then it is called a ground term.

Definition 2.6 (Atomic Formula)**.** *An* atomic formula *is an n-ary predicate letter followed by an expression of the form* $(t_1, t_2, \ldots, t_n)$*, where each* t_i *is a term.*

Definition 2.7 (Well-formed formulas)**.** Well-formed formulas *(*wff*'s) are defined inductively.*

1. *Each atomic formula is a* wff.
2. *If* $\mathbf{P}$ *is a* wff*, then* $(\neg\mathbf{P})$ *is a* wff.
3. *If* $\mathbf{P}$ *and* $\mathbf{Q}$ *are* wff*'s, then* $(\mathbf{P} \rightarrow \mathbf{Q})$ *and* $(\mathbf{P} \wedge \mathbf{Q})$ *are* wff*'s.*
4. *If* $\mathbf{P}$ *is a* wff *and* $\mathbf{x}$ *a variable, then* $\forall\mathbf{x}\mathbf{P}$ *and* $\exists\mathbf{x}\mathbf{P}$ *are* wff*'s.*
5. *Nothing else is a formula.*

A formula of the form $(P \rightarrow Q)$ is also called a *rule*. P is the *antecedent* of the rule, while Q is the *consequent*. Alternatively, P is the *body* of the rule and Q is the *head*. Logic programs consist of sets of a restricted sort or rule, described in section 2.2.

Definition 2.8 (Free variables)**.** *Free variables in formulas are defined inductively.*

1. $\mathbf{x}$ occurs free *in* $\mathbf{P}$ *if* $\mathbf{P}$ *is an atomic formula and* $\mathbf{x}$ *occurs in* $\mathbf{P}$.
2. $\mathbf{x}$ *is free in* $(\neg\mathbf{P})$ *if* $\mathbf{x}$ *is free in* $\mathbf{P}$.
3. $\mathbf{x}$ *is free in* $(\mathbf{P} \rightarrow \mathbf{Q})$ *and* $(\mathbf{P} \wedge \mathbf{Q})$ *if* $\mathbf{x}$ *is free in* $\mathbf{P}$ *or* $\mathbf{Q}$.
4. $\mathbf{x}$ *is free in* $\forall\mathbf{v}\mathbf{P}$ *if* $\mathbf{x}$ *is free in* $\mathbf{P}$ *and* $\mathbf{x} \neq \mathbf{v}$.
5. $\mathbf{x}$ *is free in* $\exists\mathbf{v}\mathbf{P}$ *if* $\mathbf{x}$ *is free in* $\mathbf{P}$ *and* $\mathbf{x} \neq \mathbf{v}$.

A formula with no free variable is called *closed* (also, a *sentence*). Otherwise it is *open*.

2.1.1 Model Theory

The semantics of first order logic (and its notion of consequence) is based on two fundamental notions: *interpretation* and *model*. The components of the sentences are interpreted relative to some domain of discourse, truth is defined relative to interpretations, and consequence is defined using the notion of truth. The conclusion of an argument is a consequence of the premises if in every interpretation where the premises are true, the conclusion is true as well.

Definition 2.9. *An* interpretation *(or* structure*)* $\mathcal{I}$ *for a language* L *is composed of*

1. *a nonempty set* D*, the universe (domain) of discourse.*
2. *an assignment to each constant* $\mathbf{c}$ *of* L *an element* $\mathbf{c}^{\mathcal{I}}$ *of* D.
3. *an assignment to each n-ary predicate* $\mathbf{P}$ *of* L *an n-ary relation* $\mathbf{P}^{\mathcal{I}} \subseteq D^n$.
4. *an assignment to each n-ary function* $\mathbf{f}$ *of* L *a total n-ary function* $\mathbf{f}^{\mathcal{I}} : D^n \rightarrow D$.

Definition 2.10. *Let* V *be the set of variables of* L*. An* assignment $s : V \rightarrow D$ *is a total function from the set of variables of* L *to the universe of discourse* D*.* $s(\mathbf{v}|d)$ *is an assignment that differs from* s *only by mapping* $\mathbf{v}$ *to* d*, where* $d \in D$*. Let* T *be the set of terms of* L*. An extended assignment* $s^* : T \rightarrow D$ *is a total function defined inductively as follows.*

1. *if* $\mathbf{c}$ *is a constant, then* $s^*(\mathbf{c}) = \mathbf{c}^{\mathcal{I}}$.
2. *if* $\mathbf{v}$ *is a variable, then* $s^*(\mathbf{v}) = s(\mathbf{v})$.
3. *if* $\mathbf{f}$ *is an n-arity function symbol and* $\mathbf{t_1}, \ldots, \mathbf{t_m}$ *are terms, then* $s^*(\mathbf{f}(\mathbf{t_1}, \ldots, \mathbf{t_m})) = \mathbf{f}^{\mathcal{I}}(s^*(\mathbf{t_1}), \ldots, s^*(\mathbf{t_m}))$.

Definition 2.11 (Satisfaction)**.** *Let $\mathcal{I}$ be an interpretation, s an assignment, and $\mathbf{P}$ a well-formed formula. In the following, we write $s, \mathcal{I} \models \mathbf{P}$ to say that s satisfies $\mathbf{P}$ relative to $\mathcal{I}$. We define satisfaction inductively.*[1]

1. *If $\mathbf{P}(t_0, \ldots, t_n)$ is an atomic formula, then*

   $s, \mathcal{I} \models \mathbf{P}(\mathbf{t_0}, \ldots, \mathbf{t_n})$ *iff* $(s^*(\mathbf{t_0}), \ldots, s^*(\mathbf{t_n})) \in \mathbf{P}^{\mathcal{I}}$.
2. $s, \mathcal{I} \models (\neg \mathbf{P})$ *iff* $s, \mathcal{I} \not\models \mathbf{P}$.
3. $s, \mathcal{I} \models (\mathbf{P} \rightarrow \mathbf{Q})$ *iff* $s, \mathcal{I} \not\models \mathbf{P}$ *or* $s, \mathcal{I} \models \mathbf{Q}$.
4. $s, \mathcal{I} \models (\mathbf{P} \wedge \mathbf{Q})$ *iff* $s, \mathcal{I} \models \mathbf{P}$ *and* $s, \mathcal{I} \models \mathbf{Q}$.
5. $s, \mathcal{I} \models \forall \mathbf{x} \mathbf{P}$ *iff* $s(\mathbf{x}|d), \mathcal{I} \models \mathbf{P}$ *for all elements* $d \in D$.
6. $s, \mathcal{I} \models \exists \mathbf{x} \mathbf{P}$ *iff* $s(\mathbf{x}|d), \mathcal{I} \models \mathbf{P}$ *for some element* $d \in D$.

Definition 2.12 (Model)**.** *If $s, \mathcal{I} \models \mathbf{P}$ for all assignments s for $\mathcal{I}$, then we say that $\mathbf{P}$ is* true *in $\mathcal{I}$, and that $\mathcal{I}$ is a* model *of P. For a set of formulas S, $\mathcal{I}$ is a model of S if it is a model for each element of S.*

Consequence in first order logic is based on this notion.

Definition 2.13 (Consequence)**.** *Let $\mathbf{P}$ be a sentence of a first order language and $\mathbf{S}$ a set of such sentences. Then $\mathbf{P}$ is a consequence of $\mathbf{S}$, written $\mathbf{S} \models \mathbf{P}$, iff every model of $\mathbf{S}$ is also a model of $\mathbf{P}$.*[2]

Every sentence of FOL can be translated into a logically equivalent sentence in which the quantifiers appear only at the beginning of the sentence. Such a sentence is said to be in *prenex normal form*. Furthermore, existentially quantified variables can be removed through a process called *Skolemization*: If $\forall x_1 \forall x_2 \ldots \forall x_1 \exists y \mathbf{P}$ is a sentence in prenex normal form, then we eliminate $\exists y$ and replace each occurrence of y in $\mathbf{P}$ with $f(x_1, x_2, \ldots x_n)$, where f is an entirely new function symbol. A set of prenex normal sentences in which the existential quantifiers have been removed is said to be in *Skolem normal form*. In general, the process of Skolemization does not yield a sentence that is logically equivalent to the original. Importantly however, if $\mathbf{P}$ is a sentence and $\mathbf{P}'$ the result of Skolemization, then $\mathbf{P}$ has a model *iff* $\mathbf{P}'$ does.

[1] Enderton uses the notation $\models_{\mathcal{I}} \mathbf{P}[s]$ where we use $s, \mathcal{I} \models \mathbf{P}$.

[2] In Tarski's (translated) words: "We say that the sentence X follows logically from the sentences of the class K if and only if every model of the class K is at the same time a model of the sentence X." [Tar02]

In the following, it is assumed that all sentences are in Skolem normal form. Since this is so, it is superfluous to even write down the quantifiers (they will be considered implicitly present).

2.1.2 HERBRAND MODELS

In the above definitions, there is no restriction on what set might be taken as the universe of discourse or how the interpretation is defined. To determine whether a given set of formulas has a model, however, it is sufficient to consider only a special sort of interpretation.

Definition 2.14 (Herbrand Universe, Base, and *Lit*)**.** *Let L be a first order language. The* Herbrand universe *U_L of L is the set of ground terms that can be created from the symbols of L. If L contains no constants, then a new constant symbol appears in U_L. The* Herbrand base *B_L of L is the set of all atomic formulas that can be constructed from the Herbrand universe and the predicates of L. The set Lit_L is $B_L \cup \{\neg p | p \in B_L\}$.*

Definition 2.15 (Ground Instantiation)**.** *The* ground instantiation *of a set of open formulas S is the set of sentences obtained via any assignment of variables of S to elements of U_L.*

Lit_L is sometimes called the *extended* base. It is noted that if function symbols are allowed, then the Herbrand universe is infinite, and so the ground instantiation is infinite.

Definition 2.16 (Herbrand Interpretations)**.** *Let S be a set of sentences of a first order language L and B_L its Herbrand base. A Herbrand interpretation $\mathcal{I}$ is an interpretation in which*

1. *the domain of discourse is U_L, and*
2. *for each constant* **c** *of L, $\mathbf{c}^{\mathcal{I}} = \mathbf{c}$,*
3. *for each ground term $f(t_0, t_1, \ldots t_n)$, $f^{\mathcal{I}}(t_0, t_1, \ldots, t_n) = f(t_0, t_1, \ldots, t_n)$, and*

Let $B_L[p] = \{p(t_0, t_1, \ldots, t_n) |\ p(t_0, t_1, \ldots, t_n) \in B_L$ where $n \geq 0\}$. $B_L[p]$ is the set of ground atoms with predicate p. From the definitions, it can be seen that in a Herbrand interpretation $\mathcal{I}$, $p^{\mathcal{I}}$ will correspond to some set $S \subseteq B_L[p]$. In this way a Herbrand interpretation $\mathcal{I}$ can be viewed alternatively as a partition of B_L into 2 sets (the *true* and the *false*) or more simply as a set $X \subseteq B_L$ (where X is taken as the set of atoms true in $\mathcal{I}$, and $B_L - X$ constitutes the set of false atoms).

Because these interpretations partition B_L into two sets, we will sometimes refer to them as 2–valued interpretations. Later sections discuss interpretations that partition B_L into three or more sets.

Theorem 2.17 (Herbrand's Theorem [Her30][3])**.** *Let S be a set of sentences from a first order language L. Then S has a model iff S has a Herbrand model.*

As was stated at the beginning of this section, in order to determine whether a set S has a model or whether $S \models p$, we thus need only concern ourselves with Herbrand interpretations and models. This is important in the context of logic programming, which deals only with syntactic structures. Particularly, early logic program semantics were specified in terms of Herbrand interpretations.

2.2 Logic Programs

The fundamental syntax of logic programs largely coincides with that of a standard first order language. However, there are a few significant differences. Quantifiers are not used; all sentences containing variables are instead assumed to be in Skolem normal form. The comma ',' is used in place of $\wedge$ to denote conjunction. The material implication symbol $\rightarrow$ is replaced with $:-$, (and the semantics are *not* the same). Rules are written with the consequent to the left of the antecedent. There are two forms of negation: (1) '$\neg$' called *classical* or *explicit* or *strong* negation; and (2) 'not', called *default negation* (or *negation-as-failure*). By convention, variables are often written using upper-case roman characters (X rather than x); predicates, functions, and constants use lower-case. Wffs made from these primitives are described below.

Definition 2.18 (Classical and Default Literals)**.** *Atoms are defined as they are in FOL. A* classical literal *is either an atom or else an atom preceded with a single occurrence of the negation symbol '$\neg$'. An atom p is called a positive literal, and $\neg p$ is called a negative literal. A* default *literal is a classical literal preceded with a single occurrence of the symbol ' not '.*

[3]This is an important result in general, and is of central importance in logic programming. The theorem is stated in most logic programming textbooks, *e.g.* [Llo87], [Hog90].

Definition 2.19 (Rules)**.** *A* logic program rule *is a finite structure of the form*

$$p \text{ :- } a_1, a_2, \ldots, a_n,\ not\ b_1,\ not\ b_2, \ldots,\ not\ b_m$$

where the head p *and each* a_i *and* b_i *in the body is a classical literal and* $m, n \geq 0$*. The expressions* head(r) *and* body(r) *refer, respectively, to the head and body of rule* r*. The expression* $body(r)^+$ *refers to the non-default portion of rule* r*, whereas* $body(r)^- = \{p |\ not\ p$ *occurs in* $body(r)\}$.

A *definite* rule possesses only positive classical literals (and no default literals). A *normal* rule possesses zero or more default literals and no negative classical literals. An *extended* rule possesses zero or more negative classical literals and zero or more default literals. Historically, semantics were developed first for definite rules (using Herbrand interpretations) and only later for normal and extended rules.

Definition 2.20 (Logic Program)**.** *A* logic program *is a set of rules. If the program consists of only definite rules, then it is a* definite program*. A program consisting of normal (extended) rules is a* normal (extended) program*.*

It is incorrect to regard rules as material conditionals. Under almost all of the logic programming semantics, contraposition does not hold. The rule $p \ :- \ q$ is not equivalent to $\neg q \ :- \ \neg p$. In particular, from $\neg q$ and $q \ :- \ p$, one cannot derive $\neg p$.

In the following, we will almost exclusively restrict ourselves to propositional logics. A logical expression containing variables is taken as a schema for a countable set of ground formulas. Particularly, logic programs containing variables are taken as short-hand notation for a set of ground rules.

2.3 Functions on Complete Lattices

Many of the semantics defined for logic programs involve complete lattices of interpretations and operators upon them. The same holds true for the defeasible logic semantics defined in later chapters. It is thus necessary to recall some basic definitions from lattice theory. Also, a few theorems of particular importance are restated.

Definition 2.21 (Partial Order)**.** *Let L be a set and $\preceq$ a binary relation on L. $\langle L, \preceq \rangle$ is a* partial order *iff $\preceq$ is reflexive, antisymmetric, and transitive.*

Definition 2.22 (Complete Lattice)**.** *A partial order $\langle L, \preceq \rangle$ is a* lattice *iff every nonempty finite subset $S \subseteq L$ has a least upper bound (lub) and greatest lower bound (glb). A lattice is* complete *if* all *subsets have a lub and glb.*

Every complete lattice has a least element (denoted $\bot$) and a greatest element ($\top$). If B is a set of logical atoms and 2^B its powerset, then it is clear that $\langle 2^B, \subseteq \rangle$ forms a complete lattice with $\bot = \emptyset$ and $\top = B$. This applies to Herbrand interpretations as well, and so it can be seen that the set of all Herbrand Interpretations forms a complete lattice under set containment.

Definition 2.23 (Directed Set)**.** *A set $S \subseteq L$ is* directed *iff* $(\forall u \in S)(\forall v \in S)(\exists w \in S)(u \preceq w \ \& \ v \preceq w)$.

Definition 2.24 (Monotone Function; Continuous Function)**.** *A function f on a complete lattice L is* monotone *iff* $(\forall X \in L)(\forall Y \in L)(X \preceq Y \rightarrow f(X) \preceq f(Y))$. *$f$ is* continuous *on a complete lattice L iff for every directed set $S \subseteq L$, $f(lub(S)) = lub(\{f(u) | u \in S\})$.*

In logic programming, it is common to denote nested applications of a monotone function f using superscripts. *E.g.*, $f(f(X))$ may be written $f^2(X)$. Further notation is also used. The expression $f \uparrow n$ (read "f up n") denotes $f^n(\varnothing)$ which is the result of iterating f n times using the bottom element as the starting point. The expression $f \downarrow n$ (read "f down n") denotes $f^n(\top)$, the result of iterating f n times using the top element as the starting point. For limit ordinals α, $f \uparrow \alpha$ is defined to be the least upper bound of $f \uparrow \beta$ for all $\beta < \alpha$. Similarly, $f \downarrow \alpha$ is the greatest lower bound of $f \downarrow \beta$ for all $\beta < \alpha$. Thus, for such functions, the following transfinite sequences can be defined.

$f \uparrow 0 = \bot$	$f \downarrow 0 = \top$	
$f \uparrow \alpha + 1 = f(f \uparrow \alpha)$	$f \downarrow \alpha + 1 = f(f \downarrow \alpha)$	(for successor ordinals)
$f \uparrow \alpha = lub(\{f \uparrow \beta \mid \beta < \alpha\})$	$f \downarrow \alpha = glb(\{f \downarrow \beta \mid \beta < \alpha\})$	(for limit ordinals)

As the below theorems describe, these sequences are related to the fixpoints of continuous functions on complete lattices.

Theorem 2.25 (Knaster-Tarski Theorem [Tar55]). *Let $\langle L, \preceq \rangle$ be a complete lattice and $f : L \to L$ be a monotone function on L. Then the fixpoints of f form a complete lattice as well. Particularly, f has a least and a greatest fixpoint, denoted $lfp(f)$ and $gfp(f)$ respectively.*

Theorem 2.26 (Kleene's Recursion Theorem [Kle52]). *Let $\langle L, \preceq \rangle$ be a complete lattice and $f : L \to L$ a continuous function on L. Then $lfp(f) = f \uparrow \omega$.*

If a complete lattice happens to be finite, then every monotone function over it is continuous.

2.4 Semantics of Definite Programs

Logic programs have at least one piece of unusual logical furniture—default negation—and how to interpret them has traditionally been a matter of considerable debate. In the case of definite programs, however, there is agreement. For these programs, since negation does not occur, one may define consequence as one does for FOL in general, treating the neck of the rule as material implication.

Definition 2.27 (Models of Definite Programs). *Let Π be a definite logic program. A Herbrand interpretation $\mathcal{I}$ is a* model *for Π iff for all rules $r \in \Pi$, if $body(r) \subseteq \mathcal{I}$ then $head(r) \in \mathcal{I}$.*

It is clear that the above definition satisfies the traditional definition of model for sets of conditionals. For definite programs, Herbrand models are closed under intersection (this does not hold for normal and extended programs, or for sets of logical formulas in general[4])—if $M_0, \ldots, M_n$ are Herbrand models of S, then $\bigcap_{i=0}^{n} M_i$ is also a Herbrand model of S—and so there is a unique minimal Herbrand Model (See for instance [vEK76][Hog90]). In addition, there is a relationship between this unique minimal model and the set of atomic consequences of the program when each rule $p \ :\!-\ A$ is interpreted as $A \to p$.

[4] the formula $p \vee q$ has two models, $\{p\}$ and $\{q\}$; their intersection is not a model.

Theorem 2.28 ([vEK76])**.** *Let* Π *be a definite program,* M_Π *its minimal Herbrand model, and* p *an atom of* B_L*. Then* $\Pi \models p$ *iff* $p \in M_\Pi$.

The minimal Herbrand model defines the consequences of Π. It is further shown in [vEK76] that this model is the least fixpoint of a continuous operator over the lattice of Herbrand interpretations. It is this latter mode of presenting the minimal model that is commonly used in logic programming.

Definition 2.29 (Immediate Consequence Operator)**.** *Let* Π *be a ground definite logic program and* $\mathcal{I}$ *a Herbrand interpretation.* $T_\Pi(\mathcal{I})$ *is defined as*

$$T_\Pi(\mathcal{I}) = \{head(r) |\ r \in \Pi \textit{ and } body(r) \subseteq \mathcal{I}\}.$$

The operator is continuous on the lattice $\langle 2^{B_\Pi}, \subseteq \rangle$ and defines the sequence

1. $T_\Pi \uparrow 0 = \varnothing$
2. $T_\Pi \uparrow \alpha + 1 = T_\Pi(T_\Pi \uparrow \alpha)$ (for successor ordinals)
3. $T_\Pi \uparrow \alpha = \bigcup_{\beta<\alpha} T_\Pi \uparrow \beta$ (for limit ordinals)

For a given definite program Π, since T_Π is continuous on $\langle 2^{B_\Pi} \rangle$, by Kleene's Theorem it follows that the least fixpoint is $T_P \uparrow \omega$. We will define $Cl(\Pi)$ to be $T_\Pi \uparrow \omega$.

Theorem 2.30. *[vEK76] Let* Π *be a ground definite program and* M_Π *its unique minimal Herbrand model. Then* $M_\Pi = T_\Pi \uparrow \omega$.

2.5 Stable Model and Answer Set Semantics

Only definite programs are guaranteed to have unique minimal Herbrand models. Consider the extended program below.

Example 2.31.

1. $p \text{ :- } \neg q$.

If the single rule is again interpreted a material conditional, then it has the two minimal models $\{p, \neg q\}$ and $\{q, \neg p\}$ and their intersection is not itself a model. Programs with default negation are in an even worse position, since without a formal interpretation of default negation it is not even clear what 'model' means for rules involving them.

Historically there has been considerable debate over what to count as the consequences of normal and extended programs. Attempts to define semantics for normal logic programs predate those for extended programs, and it is often the case that the semantics for extended programs are merely straightforward extensions of those for normal programs. For normal programs, the *stable model semantics* [GL88] is the dominant semantics used (as of 2006, Gelfond and Lifschitz's original presentation of the semantics [GL88] is the 19^{th} most cited article in the CiteSeer index [Cit06]). The general idea is to provide an interpretation for normal programs by reducing them in some sense to definite programs and then use the semantics already defined for definite programs.

Definition 2.32 (Program 'Reduct' [GL88])**.** *Let* Π *be a normal logic program and* $\mathcal{I}$ *a Herbrand interpretation. The* reduct *of* Π wrt $\mathcal{I}$*, written* $\Pi^{\mathcal{I}}$ *is the program obtained by the following process:*

1. *Delete from* Π *all rules* r *such that* $body(r)^- \cap \mathcal{I} \neq \varnothing$.
2. *From all remaining rules, delete all default literals.*

The following passage provides an informal justification for the reduct:

> If S is the set of ground literals that the agent believes to be true, then any rule that has a subgoal $not\ L$ with $L \in S$ will be of no use to him, and he will view any subgoal $not\ L$ with $L \notin S$ as trivial. Thus he will be able to replace the set of rules Π by the simplified set of rules Π^S. [GL91]

Note that the reduct of a program is a definite program and thus has a least Herbrand model. This is used to define an operator γ_Π, usually called the Gelfond-Lifschitz operator. This operator is used to define the "stable models" of a program.

Definition 2.33 (GL-Operator [GL88])**.** *Let* Π *be a normal logic program and* $\mathcal{I}$ *a Herbrand interpretation.* $\gamma_\Pi(\mathcal{I}) = Cl(\Pi^{\mathcal{I}})$.

Definition 2.34 (Stable Model [GL88])**.** *Let* Π *be a normal logic program and* $\mathcal{I}$ *a Herbrand interpretation.* $\mathcal{I}$ *is a* stable model *of* Π *iff* $\mathcal{I} = \gamma_\Pi(\mathcal{I})$.

A stable model amounts roughly to a minimal set of consistent beliefs that conforms to the rules of the program. They constitute the beliefs a "rational agent" might espouse *wrt* the program. Thus stable models are quite similar to extensions in default logic or stable expansions in autoepistemic logic (and in fact close formal ties between the formalisms have been shown). For definite programs, there is a unique stable model—the unique minimal Herbrand model. Normal programs might have zero, one, or many stable models. For instance, the first example below has no stable models, while the second has two.

Example 2.35.

1. $p \ :- \ not\ p.$

Example 2.36.

1. $p \ :- \ not\ q.$

2. $q \ :- \ not\ p.$

In the first, the only possible interpretations are $\{\}$ and $\{p\}$. If $\mathcal{I} = \{\}$, then $\Pi^\mathcal{I} = \{p\}$ (the body of the rule has been removed) and $Cl(\Pi^\mathcal{I}) = \{p\}$. If $\mathcal{I} = \{p\}$, then $\Pi^\mathcal{I} = \{\}$ (the entire rule has been removed) and $Cl(\Pi^\mathcal{I}) = \{\}$. Thus, neither interpretation is stable. In the second example, the possible interpretations are $\{\}$, $\{p\}$, $\{q\}$, and $\{p, q\}$. Only the middle two are stable.

ANSWER SET SEMANTICS

The *answer-set semantics* [GL91], generalizes the stable model semantics to extended logic programs. The name 'answer-set' is used to accentuate that the stable sets obtained can be viewed as alternative answers to a given problem. The semantics also goes by the abbreviation ASP (*Answer Set Programming*). The reduct of a program and its closure are computed in almost precisely the same way. However, since classical negation is allowed and in particular can appear in the heads of rules, it is possible for $Cl(\Pi^\mathcal{I})$ to contain complementary literals. By design, the answer-set

semantics is *explosive*; if $Cl(\Pi^{\mathcal{I}})$ contains complementary literals, then the consequences of $\Pi^{\mathcal{I}}$ are taken to be all of Lit. The operator Cn formalizes this.

Definition 2.37. *Let* Π *be an extended program containing no default literals. Define* $Cn(\Pi)$ *to be* $Cl(\Pi)$ *if* $Cl(\Pi)$ *is consistent; otherwise* $Cn(\Pi) = Lit_{\Pi}$.

In computing the reduct of an extended program, negative classical literals are treated simply as new atoms. I.e., the literal $\neg p$ has nothing at all to do with p. As this is so, 2-valued interpretations, which do not contain classically negated literals, no longer suffice. Instead, 3 and 4-valued interpretations are used.

Definition 2.38 (Consistent set)**.** *A set* S *of literals is* consistent *iff no pair of complementary literals appears in* S.

Definition 2.39 (3-valued interpretation)**.** *Let* Π *be an extended logic program. A 3-valued interpretation* $\mathcal{I}$ *of* Π *is any consistent subset of* Lit_{Π}. *An atom* p *is* true *in* $\mathcal{I}$ *iff* $p \in \mathcal{I}$. *An atom* p *is* false *in* $\mathcal{I}$ *iff* $\neg p \in \mathcal{I}$. p *is* undefined *otherwise.* $\neg p$ *is false in* $\mathcal{I}$ *if* p *is true in* $\mathcal{I}$, *and* $\neg p$ *is true if* p *is false in* $\mathcal{I}$.

Since not every atom is true or false in a given $\mathcal{I}$, $\mathcal{I}$ can also be called a *partial interpretation*.

Definition 2.40 (4-valued interpretation)**.** *Let* Π *be an extended logic program. A 4-valued interpretation* $\mathcal{I}$ *of* Π *is any subset of* Lit_{Π}. *Truth and falsity are defined as for 3-valued interpretations. A literal is undefined if it is neither true nor false.*

Observe that in a 4–valued interpretation, a given literal can be both true *and* false, or neither true nor false. Using 3-valued interpretations and Cn in place of Cl in the definition of γ_{Π}, it is possible to provide semantics for many extended programs.

Example 2.41.

1. $flies \ :- \ not \ \neg flies, bird.$
2. $\neg flies \ :- \ not \ flies, penguin.$
3. $bird.$

4. $penguin$.

In this example, the stable sets are $\{bird, penguin, flies\}$ and $\{bird, penguin, \neg flies\}$ which intuitively are the states of affairs that might obtain. However, consider the program below.

Example 2.42.

1. p.
2. $\neg p$.

While this program has no consistent stable sets, the set $\{p, \neg p\}$ *is* stable and under the answer-set semantics, p can be considered both true and false. The answer set semantics is defined using 4-valued interpretations.

Definition 2.43 (Answer Set [GL91])**.** *Let* Π *be an extended logic program and* $\mathcal{I}$ *a subset of* Lit_Π. *Define* $\gamma_\Pi(\mathcal{I}) = Cn(\Pi^{\mathcal{I}})$. $\mathcal{I}$ *is an* answer-set *of* Π *iff* $\mathcal{I} = \gamma_\Pi(\mathcal{I})$.

For Examples 2.35 and 2.36, the answer-sets and stable models coincide. Indeed, it can be seen from the definition of γ_Π for extended programs that the answer-set semantics is a strict generalization of the stable model semantics. For normal programs, the two semantics agree.

2.5.1 Consequences in SM/ASP Semantics

Consequence in the SM and ASP semantics is indirectly skeptical (in the sense of Chapter 1). The consequence relations are defined the same way for each, save that for the SM semantics 2-valued interpretations are used, while for the ASP semantics 4-valued interpretations are used. The below presentation is taken from [Bar03].

Definition 2.44 (Consequence for Stable Model Semantics)**.** *Let* Π *be a normal program and* p *an atom.* $\Pi \mid\!\approx_{SM} p$ *iff* p *is true in all 2-valued stable models of* Π. $\Pi \mid\!\approx_{SM} \neg p$ *iff* p *is false in all 2-valued stable models. Alternatively,* $\Pi \mid\!\approx_{SM} p$ *iff for each 2-valued stable model* S *of* Π, $p \in S$. $\Pi \mid\!\approx_{SM} \neg p$ *iff for all 2-valued stable models* S *of* Π, $p \notin S$.

Definition 2.45 (Consequence for Answer Set Semantics)**.** *Let* Π *be an extended program and* p *a literal.* $\Pi \mid\!\approx_{AS} p$ *iff* p *is true in all answer sets of* Π. $\Pi \mid\!\approx_{AS} \neg p$ *iff* p *is false in all answer sets of* Π.

In Example 2.41, the intersection of the two answer sets is simply $\{bird, penguin\}$, and in that example this is the intuitively correct skeptical result. There is no reason to prefer the conclusion $flies$ to $\neg flies$, and so the skeptic should choose to accept neither.

The below example, taken from [Dix94], shows that the stable model semantics (and, by extension, the answer-set semantics) fails cautious monotony.

Example 2.46.

1. a :– $not\ b$.
2. b :– $not\ a$.
3. c :– $not\ a$.
4. c :– $not\ c$.

Here $\{b, c\}$ is the unique stable model and so the set of skeptical consequences contains c. However adding c as a fact makes $\{a, c\}$ a model as well, and so b is no longer in the intersection of the models (and so no longer a consequence)

2.6 Well-Founded Semantics

The other dominant semantics for logic programs is the *well-founded semantics* (WFS) [GRS88] [GRS91]. Unlike the stable model or answer-set semantics, the WFS is directly skeptical. Furthermore, unlike under SM/ASP semantics, every program possesses exactly one well-founded model. There are several equivalent characterizations of the semantics (we have found at least eight, in fact); the original one, shown below, is found in [GRS91], though we have modified the notation somewhat here.

Originally, the WFS defined interpretations only for normal logic programs, and 3–valued interpretations were spelled out as consistent sets of literals (in other words, as we have defined them). Given a 3–valued interpretation S, we prefer to write it as $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$, where $p \in \mathcal{T}$ iff $p \in S$ and $p \in \mathcal{U}$ iff $\neg p \in S$. It is clear that $\mathcal{T}$ and $\mathcal{U}$ are disjoint.

The well-founded semantics is based upon the concept of an unfounded set of literals, which we define now.

Definition 2.47 (Unfounded Set [GRS91])**.** *Let* Π *be a ground normal program and* $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ *a 3-valued interpretation. Then* $S \subseteq B_\Pi$ *is an* unfounded *set of* Π *relative to* $\mathcal{I}$ *iff for each* $p \in S$*, one of the following holds for every rule of* Π *with head* p*:*

1. *There is a* $q \in body(r)^+$ *such that* $q \in \mathcal{U} \cup S$.
2. *There is a* $not\ q$ *such that* $q \in body(r)^-$ *and* $q \in \mathcal{T}$.

Example 2.48.

1. $r.$
2. $p\ :\!-\ p.$
3. $s\ :\!-\ p.$
4. $q\ :\!-\ not\ q.$

If $\mathcal{I} = \langle \varnothing, \varnothing \rangle$, then $\{p, s\}$ is an unfounded set *wrt* $\mathcal{I}$, as is $\{p\}$ and $\{\}$, but no other set is. Unfounded sets are used to prevent circular reasoning from justifying the consequent of a rule. In the above, there is no way to derive p without first deriving p, and so $\{p\}$ is unfounded regardless of the interpretation used. This set can be expanded to $\{p, s\}$ since the only rule for s depends upon p. q is not unfounded according to the above definition because, while it depends upon itself, it does so through default negation.

Unfounded sets are closed under union, and so for any Π and $\mathcal{I}$, there exists a *greatest unfounded set* of Π *wrt* $\mathcal{I}$, denoted $U_\Pi(\mathcal{I})$:

Definition 2.49. $U_\Pi(\mathcal{I}) = \bigcup \{A \mid A \text{ is an unfounded set of } \Pi \text{ with respect to } \mathcal{I}\}$.

$U_\Pi(\mathcal{I})$ can be viewed as a monotone operator and is used to derive the 'negative' consequences of a program. The 'positive' consequences of the program are defined by the *immediate consequence operator* T, suitably modified to encompass normal rules.

Definition 2.50 (Immediate Consequence for Normal Programs)**.**

$$T_\Pi(\mathcal{I}) = \{p \mid r \in \Pi,\ head(r) = p,\ body(r)^+ \subseteq \mathcal{T} \text{, and } body(r)^- \subseteq \mathcal{U}\}.$$

These two operators are combined to form a third, which in turn is used to define the well-founded model of the program:

Definition 2.51 (Well-Founded Operator [GRS91]). $W_\Pi(\mathcal{I}) = \langle T_\Pi(\mathcal{I}), U_\Pi(\mathcal{I})\rangle$

$W_\Pi(\mathcal{I})$ is monotonic and is used to define the sequence

1. $\mathcal{I}_0 = W_\Pi \uparrow 0 = \bot$
2. $\mathcal{I}_{\alpha+1} = W_\Pi \uparrow \alpha + 1 = W_\Pi(\mathcal{I}_\alpha)$ (for successor ordinals)
3. $\mathcal{I}_\alpha = W_\Pi \uparrow \alpha = \langle \bigcup_{\beta<\alpha} \mathcal{T}_\beta, \bigcup_{\beta<\alpha} \mathcal{U}_\beta\rangle$ (for limit ordinals)

Definition 2.52 (Well-Founded Model [GRS91]). *Let* Π *be a normal logic program. The well-founded model of* Π*, written* $wfm(\Pi)$*, is the least fixpoint of* W_Π*.*

The operator W_Π is not continuous in general but since it is monotone, for finite programs $W_\Pi \uparrow \omega = lfp(W_\Pi)$. The well-founded model of Example 2.48 is $\langle\{r\}, \{p, s\}\rangle$. q is not unfounded since there is no $\mathcal{I}_\alpha$ in which q appears in $\mathcal{U}_{\mathcal{I}_\alpha}$. Simply speaking, according to the definition of unfounded sets, in order to make q unfounded *wrt* $\mathcal{I}_\alpha$, we must first make it unfounded *wrt* $\mathcal{I}_\beta$ for some $\beta < \alpha$, and this we cannot do.

The well-founded semantics divides sets of literals into 3 sets: the well-founded, the unfounded, and the neither. We can use this to define two consequence relations (or one consequence relation and one *anti-consequence* relation).

Definition 2.53 (Well-Founded Consequences). *Let* Π *a normal logic program and* $\mathcal{I} = \langle \mathcal{T}, \mathcal{U}\rangle$ *its well-founded model. Then* $\Pi \mid\!\approx_{WFS} p$ *iff* $p \in \mathcal{T}$*, and* $\Pi \approx\!\mid_{WFS} p$ *iff* $p \in \mathcal{U}$*.*

Importantly, the well–founded semantics *approximates* the stable model semantics, in that the positive portion of the well–founded model is contained in the intersection of all stable models, while the negative portion is contained in no stable model. In the case of definite programs, the positive portion of the well-founded model corresponds exactly to the unique stable model of the program, which is also the minimal Herbrand model.

Proposition 2.54. *Let* Π *be a ground normal program,* $wfm(\Pi) = \langle \mathcal{T}, \mathcal{U}\rangle$*, and* p *an atom. If* $p \in \mathcal{T}$*, then* p *is true in every stable model. If* $p \in \mathcal{U}$*, then* p *is false in every stable model.*

2.6.1 THE ALTERNATING FIXPOINT PROCEDURE FOR THE WFS

By a happy coincidence, the Gelfond-Lifschitz operator γ defined for the stable model and answer-set semantics can also be used to formulate an alternative characterization of the well-founded semantics [BS93] [Gel93]. As is evident from its definition, the γ operator is antimonotone.

Proposition 2.55 (γ is antimonotone)**.** *Let Π be a logic program and S and T sets of atoms. If $S \subseteq T$ then $\gamma_\Pi(T) \subseteq \gamma_\Pi(S)$.*

As γ is antimonotone, γ^2 is monotone. It turns out that the well-founded model can be obtained from the least fixpoint of this monotone operator.

Theorem 2.56 ([BS93])**.** *Let Π be a normal logic program and $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ its well-founded model. Then $p \in \mathcal{T}$ iff $p \in lfp(\gamma^2)$, and $p \in \mathcal{U}$ iff $p \notin gfp(\gamma^2)$.*

The well-founded model can be approached from below by iterating γ^2 on $\bot$. Informally, if we start with $S_0 = \varnothing$, then $\gamma_\Pi(S)$ yields a set S_1 of all the atoms that are *potentially* well-founded (and correspondingly, $Lit_\Pi - \gamma_\Pi(S)$ is a set of literals known to be unfounded). That is, if $wfm(\Pi) = \langle \mathcal{T}, \mathcal{U} \rangle$, then $\gamma_\Pi(S_0)$ is an *overestimate* $\mathcal{T}$ and $Lit_\Pi - \gamma_\Pi(S)$ is an underestimate of $\mathcal{U}$. A second application of the operator yields an *underestimate* S_2 of $\mathcal{T}$ (and an underestimate of $\mathcal{U}$). With each double application of the operator, better and better approximations of $\mathcal{T}$ and $\mathcal{U}$ are obtained.

Example 2.57.

1. a :– b.
2. a :– c.
3. b :– $not\ c$.
4. c :– $not\ b$.
5. d :– e.
6. e :– $not\ f$.

The above example has the stable models $\{a, b, d, e\}$ and $\{a, c, d, e\}$, and their intersection is $\{a, d, e\}$. Note that f appears in no model. The well-founded model is $\langle \{d, e\}, \{f\} \rangle$. Neither a nor b is true or false in the well-founded model. Observe that a is a consequence of the program under the stable model semantics but not under the WFS.

2.6.2 THE COHERENCE PRINCIPLE

The WFS can be applied to extended logic programs without modification (see [BG94], [Prz91] [Lif96], or [Bre96]). One may treat negative literals simply as new atoms. This effectively yields a normal logic program, and the well-founded model can be computed in the usual fashion. However, since negative literals are used, inconsistent models are possible. When this occurs, one might simply say that this is not a model of the program (as in [Prz91]), or say that the well-founded literals are all of Lit_Π (as in [BG94]). The latter is the natural result of using γ_Π^2 to define the model, where $\gamma_\Pi(\mathcal{I}) = Cn(\Pi^\mathcal{I})$.

Theorem 2.58 (WFS for Extended Logic Programs [BG94])**.** *Let* Π *be an extended logic program. Then* p *is* true wrt *the well-founded semantics of* Π *iff* $p \in lfp(\gamma^2)$*, and* p *is* false *iff* $p \notin gfp(\gamma^2)$*. Otherwise* p *is undefined.*

However, defining the WFS for extended programs in this fashion does not always lead to results that are acceptable to everyone.

Example 2.59. *(from [PA92])*

1. $a \ :- \ not\ b.$
2. $b \ :- \ not\ a.$
3. $\neg a.$

The well-founded model of the above program is $\langle\{\neg a\}, \{\}\rangle$. The literals a and b are undefined while $\neg a \in \mathcal{T}$. Alferes and Pereira find this result puzzling and in [PA92], [ADP95],[AP96] recommend imposing a 'coherence restriction' upon the program. Specifically, for any literal p, if $p \in \mathcal{T}$ then $\neg p \in \mathcal{U}$. This yields a new semantics called the $WFSX$. It is found that the coherence restriction can be met by modifying the γ operator.

Definition 2.60 (Semi-Normal Form [ADP95])**.** *If* Π *is a logic program and* r

$$p \ :- \ a_1, \ldots a_n, \ not\ b_1, \ldots b_m$$

is a rule in Π*, then the rule* r_s

$$p \ :- \ a_1, \ldots a_n, \ not\ b_1, \ldots b_m, \ not\ \neg p$$

is the semi-normal *form of* r. *The* semi-normal *form of* Π, *written* Π_s *is* $\{r_s |\ r \in \Pi\}$.

Definition 2.61 (Partial Stable Model [ADP95]). *Let* Π *be an extended logic program and* $\mathcal{T}$ *and* $\mathcal{U}$ *disjoint sets of literals.* $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ *is a partial stable model of* Π *iff:*

1. $\mathcal{T} = \gamma_\Pi(\gamma_{\Pi_s}(\mathcal{T}))$
2. $\mathcal{T} \subseteq \gamma_{\Pi_s}(\mathcal{T})$
3. $\mathcal{U} = Lit_\Pi - \gamma_{\Pi_s}(\mathcal{T})$

Definition 2.62 (Contradictory Program [ADP95]). *Let* Π *be an extended logic program.* Π *is* contradictory *if for all answer sets* X, *there exists an atom* p *such that* $p \in X$ *and* $\neg p \in X$.

Definition 2.63 (WFSX [ADP95]). *If* Π *is a noncontradictory extended logic program, then under the WFSX,* $wfm(\Pi) = \langle lfp(\gamma_\Pi \gamma_{\Pi_s}), Lit_\Pi - \gamma_{\Pi_s}(\mathcal{T}) \rangle$.

Note that the second condition in Definition 2.61 above is intended to prevent contradictory interpretations. If both a and $\neg a$ appear in X, then all rules for a and $\neg a$ will be deleted in Π_s^X, and so neither literal can appear in γ_{Π_s}. Dropping this condition yields a paraconsistent version of the semantics, called $WFSX_p$ [Dam96][DP97].

2.7 Remarks on Complexity

As was stated, the stable model/answer-set semantics and the well-founded semantics constitute the dominant semantics for logic programs today. In later chapters we will explore the relationship between these semantics and defeasible logic. We end this chapter with a discussion of certain complexity issues that are relevant, and mention a few implementations of logic programming under these semantics that exist.

It is well-known that determining satisfiability of FOL formulas is undecidable. Similarly, given an arbitrary definite program Γ with function symbols and a ground atom p, determining whether $\Gamma \models p$ is undecidable. It is, however, recursively enumerable, and for ground definite programs, the least Herbrand model can be computed in linear time *wrt* to the size of the program [DG84].

For a definite program with function symbols, since the stable model semantics and WFS coincide with the unique minimal Herbrand model, computing membership of a ground atom

in the well-founded model or unique stable model is recursively enumerable. For normal and extended programs, the problem is not; this is proven in the next section. For ground normal and extended programs, however, the WFM can be computed via the alternating fixpoint procedure in quadratic time relative to the size of the program. Specifically, it can be done in time $O(|\Pi||A|)$ where $|\Pi|$ is the size of the program and $|A|$ is the number of distinct literals appearing in Π [GRS91][Wit90]. An alternative transformation-based procedure for computing the WFM is given in [ZFB96]; the worst-case performance does not improve upon the alternating fixpoint procedure, however. According to [LT00], no known algorithm performs under the quadratic barrier; it is an open problem whether linear time computation of the WFM is possible.

Deciding whether a finite ground normal program has a stable model is NP-Complete [MT91]. Given a set $\mathcal{I}$, creating $\Pi^{\mathcal{I}}$ can be done in polynomial time, as can computing the least Herbrand model of $\Pi^{\mathcal{I}}$, and so checking whether $\mathcal{I}$ is a stable model can be done in polynomial time. A 3CNF formula Φ can be readily embedded into a normal logic program Π_Φ; Φ is satisfiable if and only if Π_Φ has a stable model. Thus, 3SAT can be reduced to the problem of finding a stable model of a normal program. In the worst case, the only manner to determine whether a program has a stable model is to "guess and check".

Several implementations exist for logic programming under the semantics discussed above. The system `smodels` is a well–known system which computes the stable and well-founded models of ground logic programs [NS96]. Programs with variables (but no function symbols) are first made ground with a front end compiler called `lparse`. DLV is another implementation that computes both answer sets and WFS, though it also allows disjunctions to appear in the heads of rules [EFK$^+$00]. Both systems compute the models in a bottom-up fashion; neither allows arbitrary function symbols. `XSB` is a popular Prolog implementation that can also answer (non-floundering) queries according to the WFS [SSW94]. In contrast to the other systems, it is top-down and does allow function symbols; the entire model is not computed. Other systems are `ASSAP` [LZ02], and `cmodels` [GLM04] [BL03];

2.7.1 THE WFS IS NOT RECURSIVELY ENUMERABLE

Theorem 2.64. *Let Φ be a finite non-ground normal program and $\mathcal{I}_{\Phi,WF}$ its well-founded model. The language $L_\Phi = \{p | p \in \mathcal{T}_{\Phi,WF}\}$, is not recursively enumerable.*

Proof. Recall that the well-founded model consists entirely of ground literals. We note that in the case of a definite program Π, L_Π corresponds to the minimal Herbrand model of Π. Let Π be a finite non-ground definite program. It is known that the minimal model M_Π of Π is recursively enumerable but not recursive. Since it is recursively enumerable, there is a Turing machine M that halts in the accept state on $\langle p, \Pi \rangle$ when $p \in L_\Pi$. Observe that in the input $\langle p, \Pi \rangle$, p is necessarily ground but Π need not be. Let Φ be the following normal logic program.

$$\Phi = \Pi \cup \{head(r)^* \ :- \ not\ head(r). | r \in \Pi\}$$

We have simply added new rules to Π. Each $head(r)^*$ is a new atom, and is intended to serve as a witness of the *non-provability* of $head(r)$. In the translation, both $head(r)$ and $head(r)^*$ are allowed to be non-ground. It is clear that since Π is finite, Φ can be constructed in finite time. Note that each $head(r)^*$ does not appear in the definite part of Φ (that is, it is not part of Π) and so does not affect the computation of L_Π.

Suppose for a proof by contradiction that L_Φ is recursively enumerable. Then there exists a Turing machine N that halts in the accept state on $\langle p, \Phi \rangle$ when $p \in L_\Phi$. However, for any ground atom p^*, by definition of the well-founded model of Φ, $p^* \in \mathcal{T}_{\Phi,WF}$ *iff* $p \in \mathcal{U}_{\Phi,WF}$. Since the well-founded model is coherent, $p \in \mathcal{U}_{\Phi,WF}$ implies $p \notin \mathcal{T}_{\Phi,WF}$.

We may now decide L_Π by running N and M concurrently. If M accepts $\langle p, \Pi \rangle$ then $p \in L_\Pi$. If N accepts $\langle p^*, \Phi \rangle$ them $p \notin L_\Pi$. However, L_Π is not decidable. This is a contradiction, and so L_Φ is not recursively enumerable. □

For programs involving function symbols and variables, determining whether a set S is a stable model is also not recursively enumerable. For a definite program, the only stable model S is the minimal Herbrand model. In general, we can recognize when an atom is in this set, but we cannot recognize when it is not.

CHAPTER 3

DEFEASIBLE LOGIC

The earliest systems of defeasible logic appeared in 1986-87 [Nut86][Nut87]. The primary system at this time was called LDR; this system also appears in 1988's [Nut88]. The logics were designed from the start to be implemented. In structure they are similar to logic programs, and in fact the first systems were implemented in Prolog [Nut88][CNV97]. The most recent logic published by Nute himself, which we call NDL, is found in [Nut01]. This logic is ambiguity blocking. An ambiguity propagating counterpart to NDL, which we call ADL, is presented in [MN06a]. In 1993, David Billington presented a quantified version of one of Nute's logics and showed it to be cumulative [Bil93]. Billington would later go on to publish dozens of papers on this logic with Grigoris Antoniou, Michael Maher, Guido Governatori, and others. We'll call their basic logic BDL. The papers present different semantics for BDL, show relationships with other NMR formalisms, describe implementations of it as well as applications (potential and actual), and discuss problems amenable to solution by defeasible logic.

Thus there are really two main varieties of defeasible logic appearing in the literature today—the *northern variants* (NDL and ADL) and the *southern variants* (BDL and its kin). 'Northern' and 'Southern' are used simply because of the hemispheres in which the respective logics were developed. Obviously, the two camps share much in common—the language is basically the same and the logics are motivated by the same ideas. However, there are significant differences between them (some differences at least are due to differences in opinions and interests). The differences are evident in both the proof systems and semantics for the logics.

This chapter serves as a brief introduction to defeasible logic. It presents the basic syntax, discusses the primary proof systems that have been developed, and draws some relationships between

them. Each of the defeasible logics discussed has its virtues. The southern variants have low complexity and thus are more amenable to embedded applications.[1] For theories involving cycles and indirect conflicts, the northern variants are in our opinion able to draw intuitively more reasonable results. However, it should be noted that each of the logics also produces unacceptable conclusions in at least some cases.

The logic ADL is relatively new, and its proof system is presented here. Semantics for it and NDL are presented in the next chapter. In later chapters, a correlation is shown between the consequences of ADL and the well-founded semantics for logic programs.

3.1 Basic Concepts

Like logic programming, the basic expressions of defeasible logic are rules. The bodies of rules are finite sets of literals and the heads are single literals. Unlike logic programs, however, rules in defeasible logic come in three flavors: *strict* (written $A \rightarrow p$), *defeasible* ($A \Rightarrow p$), and *undercutting defeaters* ($A \rightsquigarrow p$). Furthermore, there is no such thing as default negation in defeasible logic. Sets of literals called *conflict sets* specify a conflict relationship between rules. A collection of rules conflict if their heads form a conflict set. A *priority relation* over rules provides a mechanism for resolving conflicts between rules. The distinct logics specify different constraints on each of these components and determine how they are used.

Atoms and literals are defined in defeasible logic in the same way as in logic programming. In the following discussion we restrict ourselves to propositional logics, and so all literals contain only ground terms. We will use $\dashrightarrow$ to stand for any rule, whether strict, defeasible, or defeater.

Definition 3.1 (Rules in Defeasible Logic)**.** *Let A be a finite set of literals and p a literal. Then $A \rightarrow p$ is a* strict *rule, $A \Rightarrow p$ is a* defeasible *rule, and $A \rightsquigarrow p$ is a* defeater.

When $body(r)$ is the empty set or a singleton, we will often omit the braces (writing $\Rightarrow p$ instead of $\{\} \Rightarrow p$). Defeasible theories, the analogs of logic programs, consist of sets of rules.

[1] This has in fact been done. See [McD05]

Definition 3.2 (Defeasible Theory)**.** *A* defeasible theory D *is a triple* $\langle R, C, \prec \rangle$*, where* R *is a countable set of rules, each with a possibly empty antecedent,* C *a countable set of finite sets of literals in the language of* D *such that for any literal* p *appearing in* D*,* $\{p, \neg p\} \in C$*, and* $\prec$ *an acyclic binary relation over the non-strict rules in* R*.*

Let $D = \langle R, C, \prec \rangle$ be a defeasible theory. The sets of literals in C are the *conflict sets* mentioned above. Each conflict set $c \in C$ specifies a set of literals that cannot consistently hold. If C only contains sets of the form $\{p, \neg p\}$, we say that conflict sets are minimal in D (and each such set is called a minimal conflict set). We say that conflict sets (and by extension, defeasible theories) are closed under strict rules if, for all $c \in C$, if $A \rightarrow p$ is a rule and $p \in c$, then $\{A \cup (c - \{p\})\} \in C$. We shall call a conflict set that is not minimal an *extended* conflict set. It is not a necessary condition that a defeasible theory be closed under strict rules, and indeed it is computationally very expensive to have them closed, but it is often a necessary prerequisite to drawing reasonable conclusions from a theory.

We use R_s, R_d, and R_u to denote the strict, defeasible, and defeater rules of D, respectively. The sets $R_s[p]$, $R_d[p]$, and $R_u[p]$ refer to those rules with head p. $C[p]$ denotes the set of conflict sets containing p.

The basic idea behind the proof system of each defeasible logic is that a literal p can be derived from a defeasible theory just in case p is the head of some strict or *undefeated* defeasible rule in the theory and all of the literals in the body of the rule are also derivable. A strict rule with an empty body is called a *fact* and the head of such a rule is by definition always derivable. A defeasible rule with an empty body is called a *presumption* and may be defeated by other rules. The role of defeaters is solely to defeat other arguments that might otherwise establish a literal. *E.g.*, the defeater $q \rightsquigarrow \neg p$ can be used to prevent proving p, but it cannot be used to directly prove $\neg p$. The use of undercutting defeaters may be traced back to Pollock in [Pol74]

The proof systems for the logics NDL and ADL, described below, are based directly upon argument trees composed of nodes tagged with literals. BDL and its kin use a simpler system—in those systems proofs are linear sequences of tagged literals. However, these proofs can be readily

cast as trees, and for the purposes of comparison we will do just that. Also, we will use $child(n)$ to refer to the set of children of node n, and $label(n)$ to refer to the label of n.

Definition 3.3 (Assertions)**.** *Let D be a defeasible theory and $p \in Lit_D$. The expression $+\delta_D p$ is called a* positive defeasible assertion*, while $-\delta_D p$ is called a* negative defeasible assertion*. The expression $+\Delta_D p$ is called a* positive strict assertion*, while $-\Delta_D p$ is called a* negative strict assertion*. When only a single theory is discussed, we will omit the subscripts.*

Informally, $+\Delta_D p$ and $-\Delta_D p$ mean that a demonstration (respectively, a refutation) exists for p from D using only the strict rules of D. $+\delta_D p$ and $-\delta_D p$ mean that a demonstration (refutation) exists from D when all types of rules are allowed. Note that $-\delta_D p$ is equivalent to neither $+\delta_D \neg p$ nor to a failure to prove $+\delta_D p$. Indeed, $-\delta_D p$ means that there is a demonstration that there is no defeasible proof of p from D. For $-\Delta_D p$, a similar state of affairs obtains.

A further tag, $+\Sigma$, is used by the ambiguity propagating version of BDL to indicate that a literal is supported. It is used to propagate ambiguity, *i.e.*, to record that an argument supports a literal p even though p is not derivable.

Definition 3.4 (Support)**.** *Let D be a defeasible theory and $p \in Lit_D$. The expression $+\Sigma_D p$ is called a* positive support assertion*, while $-\Sigma_D p$ is called a* negative support assertion*.*

Definition 3.5 (Defeasible Argument Tree)**.** *τ is a* defeasible argument tree *for D iff τ is a finite tree such that every node of τ is labeled with one of $+\Sigma p$, $-\Sigma p$, $+\delta p$, $-\delta p$, $+\Delta p$ or $-\Delta p$ (for some literal $p \in Lit_D$).*

In the following, we will treat the expressions "argument tree" and "defeasible argument tree" as synonymous.

Definition 3.6 (Depth)**.** *The* depth *of a node n is k iff n has k ancestors in τ. The depth of a tree is taken to be the greatest depth of any of its nodes.*

Definition 3.7 (Success and Failure)**.** *Let A be a set of literals, and n a node of a defeasible argument tree τ.*

1. A succeeds *at* n *iff for all* $q \in A$*, there is a child* m *of* n *such that* m *is labeled* $+\delta q$*.* A strictly succeeds *at* n *iff for all* $q \in A$*, there is a child* m *of* n *such that* m *is labeled* $+\Delta q$*.* A is supported *at* n *iff for all* $q \in A$ *there exits a child* m *of* n *such that* m *is labeled* $+\Sigma q$*.*
2. A fails *at* n *iff there is a* $q \in A$ *and a child* m *of* n *such that* m *is labeled* $-\delta q$*.* A strictly fails *at* n *iff there is a* $q \in A$*, and a child* m *of* n *such that* m *is labeled* $-\Delta q$*.* A *is* not supported *at* n *iff there is a* $q \in A$ *and a child* m *of* n *such that* m *is labeled* $-\Sigma q$*.*

The different defeasible logics specify different conditions that nodes in the tree must satisfy in order for the tree as a whole to constitute a proof. In part these conditions are modular—one can combine them in different ways to form a new logic. But in some cases, certain conditions make no sense if used without others. Each node in the tree bears the label of a literal adorned with one of the symbols $\pm\Sigma$, $\pm\delta$, or $\pm\Delta$; these indicate derivability or refutability of the literal.[2] A tree with root $+\Delta p$ states that p is 'definitely' derivable from the theory. A tree with root $+\delta p$ indicates that p is only 'defeasibly' derivable from the theory. This conclusion might not hold if more rules or facts are added to the theory. The corresponding negative tags, respectively, state that p is demonstrably not derivable from the theory using strict rules alone or some combination of strict and defeasible rules. In other words, it has been shown that every possible way to prove p from the theory fails. In still other words, p has been refuted.

Definition 3.8 (Derivability)**.** *Let* D *be a defeasible theory,* $p \in Lit_D$*, and* L *one of the logics defined below. Then* $D \mid\sim_L p$ *iff there is a defeasible proof* τ *in* L *such that the root of* τ *is labeled* $+\delta p$*.* $D \sim\!|_L p$ *iff there is a defeasible proof* τ *in* L *such that the root of* τ *is labeled* $-\delta p$*. If* S *is a set of literals appearing in* D*,* $D \mid\sim_L S$ *if and only if for all* $p \in S$*,* $D \mid\sim_L p$*.* $D \sim\!|_L S$ *iff for some* $p \in S$*,* $D \sim\!|_L p$*.*

The specific details of each logic's proof system are presented in the next sections. Afterwards, a comparison of the logics is made. While Billington et al. have examined earlier versions of Nute's logic and compared them to those they have since developed, an analysis *wrt* NDL (and ADL) has

[2]The use of Δ and δ appears to be an innovation of Billington. In early works, Nute simply used p^+ and to indicate 'definitely p', and $@p^+$ to indicate 'defeasibly p'. Their negative counterparts were similar. We use the innovation here for the sake of consistency of notation.

never been performed. We do that here. Not surprisingly, none of the logics are equivalent. This is true even for theories lacking cycles and priorities. In Appendix A, several examples illustrating how the logics differ are presented.

3.2 NORTHERN VARIANTS

3.2.1 THE LOGIC NDL

NDL is an ambiguity blocking logic appearing in [Don99] and [Nut01]. It is the first defeasible logic to incorporate a cycle check, thereby weeding out circular arguments. It is also the first to introduce conflict sets to deal with indirect conflicts.

Definition 3.9. *[Proof in NDL] Let D be a defeasible theory and τ an argument tree for D. τ is a defeasible proof in NDL iff for each node n of τ, one of the following obtains.*

1. $label(n) = +\delta p$ *and either*
   a. *there is an* $r \in R_s[p]$ *such that* $body(r)$ *succeeds at* n*, or*
   b. *there is an* $r \in R_d[p]$ *such that*
      i. $body(r)$ *succeeds at* n*, and*
      ii. *for all* $c \in C[p]$ *there is a* $q \in c - \{p\}$ *such that for all* $s \in R[q]$*, either* $body(s)$ *fails at* n *or else* $s \prec r$*.*
2. $label(n) = -\delta p$ *and*
   a. *for all* $r \in R_s[p]$*,* $body(r)$ *fails at* n*, and*
   b. *for all* $r \in R_d[p]$*, either*
      i. $body(r)$ *fails at* n*, or*
      ii. *there is a* $c \in C[p]$ *such that for all* $q \in c - \{p\}$*, there is a* $s \in R[q]$ *such that* $body(s)$ *succeeds at* n *and* $s \not\prec r$*.*
3. $label(n) = -\delta p$ *and* n *has an ancestor* m *in* τ *with* $label(m) = -\delta p$*, and all nodes between* n *and* m *are negative defeasible assertions.*

The third condition is called *failure-by-looping*. Since conclusions cannot be established by circular arguments, failure-by-looping can help to show that a literal cannot be derived from a defeasible theory. The advantage of adding failure-by-looping to the proof system is obvious.

Example 3.10. $D = \langle R_D, C_D, \prec_D \rangle$

R_D:	*1)*	$\{\} \rightarrow mammal$	C_D:	$\{bat, \neg bat\}$	$\prec_D$: $\emptyset$
	2)	$\{furry, has\_wings\} \Rightarrow bat$		$\{furry, \neg furry\}$	
	3)	$\{bat\} \Rightarrow furry$		$\{has\_wings, \neg has\_wings\}$	
	4)	$\{bat\} \Rightarrow has\_wings$		$\{mammal, \neg mammal\}$	
	5)	$\{bat\} \Rightarrow flies$		$\{flies, \neg flies\}$	
	6)	$\{mammal\} \Rightarrow \neg flies$			

In earlier versions of defeasible logic (including BDL below), although we could easily see that there was no way to show $D \mid\sim bat$, we could not demonstrate this in the proof theory. That is, we could not show $D \sim\!\mid bat$. Consequently, neither could we show $D \mid\sim \neg flies$. Failure-by-looping provides a mechanism for showing $D \sim\!\mid_{NDL} bat$, which then allows us to show $D \mid\sim_{NDL} \neg flies$.

In Chapter 5, when we later define a translation of defeasible theories into logic programs in such a way that the consequences of a theory (under ADL) correspond to the well-founded model for the logic program, failure-by-looping is necessary to get this correspondence. Particularly, failure-by-looping is needed to capture within the proof theory the concept of a literal being unfounded. Where the theory D above is translated into program Π_D, the literals bat, $furry$, and has_wings are all unfounded, but $\neg flies$ is well-founded. The corresponding literals are simply undetermined in versions of defeasible logic without failure-by-looping.

All defeasible logics, like the WFS for logic programs, are directly skeptical. If p is defeasibly provable, then there must be some rule for p whose body is also defeasibly provable.

Example 3.11.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{p\} \Rightarrow q$
(4) $\{\neg p\} \Rightarrow q$

The logic program analog of the above defeasible theory is

Example 3.12.

(1) $p \text{ :- } not\ \neg p.$

(2) $\neg p \ :- \ not\ p.$
(3) $q \ :- \ not\ \neg q, p.$
(4) $q \ :- \ not\ \neg q, \neg p.$

The above program has the answer sets $\{p, q\}$ and $\{\neg p, q\}$; their intersection is $\{q\}$. In NDL, both p and $\neg p$ are refuted. Since these are the only literals that support q, q itself is refuted. In the skeptical answer set semantics for logic programs, q is a floating conclusion; it belongs to every answer set even though there is no rule that supports it that belongs to every extension. Defeasible logic does not allow such floating conclusions.

3.2.2 THE LOGIC ADL

Consider the defeasible theory below in which the precedence relation is empty (from [Bre01]).

Example 3.13.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{p\} \Rightarrow \neg q$
(4) $\{\} \Rightarrow q$

We have $D \sim|_{NDL} p$, $D \sim|_{NDL} \neg p$, $D \sim|_{NDL} \neg q$, and $D \not\sim_{NDL} q$. As was said, NDL is *ambiguity blocking*. Furthermore, as was noted in Chapter 1, most researchers feel that *ambiguity propagation*, which would in this case prevent concluding q, is intuitively more reasonable. Almost all semantics for logic programs are ambiguity propagating (to our knowledge, in fact, no ambiguity blocking semantics for logic programs exists). Brewka in [Bre01] quickly dismisses an earlier version of Nute's defeasible logic precisely because it is ambiguity blocking.

It turns out that a very minor modification to the proof system of NDL produces an ambiguity propagating defeasible logic which we call ADL. The modification creating ADL affects only part 2.b.ii of Definition 3.9. It specifies that p is defeated only if every defeasible rule in support of p fails or else is defeated by a satisfied strict rule or a satisfied defeasible rule of *higher precedence* for each element $q \in c - \{p\}$ for some $c \in C[p]$. In NDL, a rule of *equal* precedence could be used. The modified proof theory is shown below.

Definition 3.14. *[Proof in ADL] τ is a defeasible proof in ADL iff τ is an argument tree for D, and for each node n of τ, one of the following obtains*

1. $label(n) = +\delta p$ *and either*
   a. *there is an* $r \in R_s[p]$ *such that* $body(r)$ *succeeds at* n*, or*
   b. *there is an* $r \in R_d[p]$ *such that*
      i. $body(r)$ *succeeds at* n*, and*
      ii. *for all* $c \in C[p]$ *there is a* $q \in c - \{p\}$ *such that for all* $s \in R[q]$*, either* $body(s)$ *fails at* n *or else* $s \prec r$*.*
2. $label(n) = -\delta p$ *and*
   a. *for all* $r \in R_s[p]$*,* $body(r)$ *fails at* n*, and*
   b. *for all* $r \in R_d[p]$*, either*
      i. $body(r)$ *fails at* n*, or*
      ii. *there is a* $c \in C[p]$ *such that for all* $q \in c - \{p\}$*, there is a* $s \in R[q]$ *such that* $body(s)$ *succeeds at* n *and* s *is strict or else* $r \prec s$*.*
3. $label(n) = -\delta p$ *and* n *has an ancestor* m *with* $label(m) = -\delta p$*, and all nodes between* n *and* m *are negative defeasible assertions.*

Apart from this modification, all other aspects of the proof system are left alone. By examining the definition of proof trees for both NDL and ADL, it can be seen that every valid proof in ADL is a valid proof in NDL.

Proposition 3.15. If D is a defeasible theory and $D \mid\sim_{ADL} p$, then $D \mid\sim_{NDL} p$. If $D \sim\!\!|_{ADL} p$, $D \sim\!\!|_{NDL} p$.

This is shown graphically in Figure 3.1. In Example 3.13, one sees that since the rules for p and $\neg p$ are of the same precedence, neither $D \sim\!\!|_{ADL} p$ nor $D \sim\!\!|_{ADL} \neg p$ can be shown (nor, for that matter, can we show $D \mid\sim_{ADL} p$ nor $D \mid\sim_{ADL} \neg p$). Because $D \sim\!\!|_{ADL} p$ cannot be shown, $D \sim\!\!|_{ADL} \neg q$ cannot be shown, and so neither can $D \mid\sim_{ADL} q$. Each of these literals is underdetermined in ADL, neither defeasibly proven nor refuted. The logic program corresponding to the above example is

(1) p :– $not\ \neg p$.

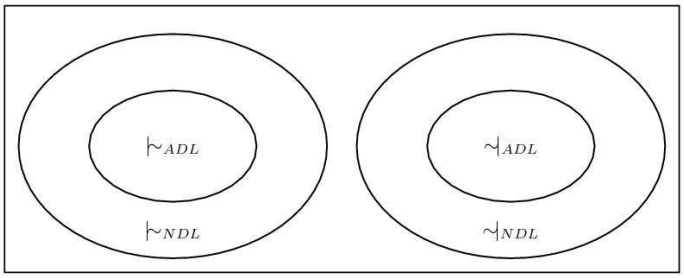

Figure 3.1: Provability in NDL and ADL.

(2) $\neg p \ :- \ not\ p.$
(3) $q \ :- \ not\ \neg q, p.$
(4) $\neg q \ :- \ not\ q.$

The answer sets of the corresponding program are $\{p, q, \}$ $\{p, \neg q\}$, and $\{\neg p, \neg q\}$. Notice that the intersection of these three extensions is empty and so the skeptical conclusions of the logic program agree with ADL in this case. Similarly, the well-founded model is $\langle \varnothing, \varnothing \rangle$.

3.2.3 The Purist's View of Strict Rules

NDL and ADL incorporate extended conflict sets to handle cases of indirect conflicts. In an example such as

(1) $\{\} \Rightarrow married$
(2) $\{married\} \rightarrow \neg bachelor$
(3) $\{\} \Rightarrow bachelor$

both NDL and ADL hold that $married$ and $bachelor$ conflict and refrain from deriving either. Antoniou calls this the "purist view" [Ant06]. It is noted that if we naively translate the theory into the logic program

(1) $married \ :- \ not\ \neg married.$
(2) $\neg bachelor \ :- \ married.$
(3) $bachelor \ :- \ not\ \neg bachelor.$

then the WFS for logic programs would conclude both $married$ and $bachelor$ to be well-founded. This is what NDL and ADL would conclude without extended conflict sets and it does not seem to us the correct conclusion for this example—quite the contrary, in fact. BDL and its variant logics do not view $married$ as conflicting with $bachelor$; but they do not go so far as to conclude $\neg bachelor$ from the above theory, either (with BDL, $married$ is provable, while both $bachelor$ and $\neg bachelor$ are refuted). Antoniou defends this view in the below passage (the literals in square braces are our own and not his):

> If p [$married$] is definitely known, then q [$\neg bachelor$] is also definitely derived. Otherwise, if p [$married$] is defeasibly known, then usually q [$\neg bachelor$]is true. In our example, p [$married$] is not definitely known, so we do not jump automatically to the conclusion q [$\neg bachelor$] once p [$married$] is (defeasibly) proven, but must consider counterarguments in favor of $\neg q$ [$bachelor$][Ant06].

We have replaced the simple literals (*e.g.*, p) with more suggestive literals (*e.g.* $married$), but we do not think this harmful, and it makes it clear that Antoniou's defense is not entirely satisfying. If $married$ is defeasibly known then it ***should*** be the case that $\neg bachelor$ is defeasibly known as well, and furthermore that $bachelor$ defeasibly refuted. We feel that the most reasonable choice in the above case is to treat strict rules as strict. How else is one to specify a linguistic convention or an analytic truth? All evidence for $married$ is evidence for $\neg bachelor$, and all evidence for $bachelor$ is evidence against $married$. Extended conflict sets were introduced in NDL and ADL to handle cases just like this.

3.3 Southern Variants

3.3.1 The Logic BDL

The proof system for the logic BDL is found in [Bil93]. It was defined for theories involving variables, but we restrict ourselves to the propositional case. We also make certain other changes (which we hope to be inessential) to make comparisons with NDL and ADL possible. Particularly,

we speak of proof trees for BDL whereas in [Bil93] proofs are simply sequences of tagged literals. Also, we treat facts as strict rules with empty bodies, which simplifies the logic.

Definition 3.16. ***(Proof in*** BDL***)*** *τ is a* defeasible proof *in BDL iff τ is an argument tree for D, and for each node n of τ, one of the following obtains.*

1. $label(n) = +\Delta p$ *and there is an* $r \in R_s[p]$ *and* $body(r)$ *strictly succeeds at* n.
2. $lable(n) = -\Delta p$ *and for all* $r \in R_s[p]$, $body(r)$ *strictly fails at* n.
3. $label(n) = +\delta p$ *and either*

   3.1 $\{p\}$ *strictly succeeds at* n, *or*

   3.1.1 *there exists an* $r \in R_{sd}[p]$ *such that* $body(r)$ *succeeds at* n, *and*
   3.1.2 $\{\neg p\}$ *strictly fails at* n, *and*
   3.1.3 *for all rules* $s \in R[\neg p]$, *either*

   3.1.3.1 $body(s)$ *fails at* n, *or*
   3.1.3.2 *there exists a* $t \in R[p]$ *such that*

   3.1.3.2.1 $body(t)$ *succeeds at* n, *and*
   3.1.3.2.2 $s \prec t$
4. $label(n) = -\delta p$ *and*

   4.1 $\{p\}$ *strictly fails at* n, *and either*

   4.1.1 *for all* $r \in R_{sd}[p]$ $body(r)$ *fails at* n, *or*
   4.1.2 $\{\neg p\}$ *strictly succeeds at* n, *or*
   4.1.3 *there exists a* $s \in R[\neg p]$*such that*

   4.1.3.1 $body(s)$ *succeeds at* n, *and*
   4.1.3.2 *for all* $t \in R[p]$, *either*

   4.1.3.2.1 $body(t)$ *fails at* n, *or*
   4.1.3.2.2 $s \not\prec t$

BDL distinguishes between strict proofs and defeasible ones, and the distinction is necessary to prevent deriving inconsistencies using strict rules that are only defeasibly supported. It is this distinction which in the above case allows BDL to conclude $married$ but prevents it from concluding $bachelor$. There is no failure-by-looping in BDL. However, given that proofs are specified to be trees, there is no reason why failure-by-looping could not be added.

The consequences of Example 3.13 according to BDL are: $-\Delta p$, $-\Delta\neg p$, $-\Delta q$, and $-\Delta\neg q$, $-\delta p$, $-\delta\neg p$, $-\delta\neg q$, $+\delta q$. It can be seen that BDL is ambiguity blocking. It is noted that BDL also incorporates what is known as *team defeat*. Consider the theory

Example 3.17.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{q\} \Rightarrow p$
(4) $\{q\} \Rightarrow \neg p$
(5) $\{\} \rightarrow q$

and suppose $2 \prec 1$ and $4 \prec 3$. In this example it is not the case that there is a single rule for p that is superior to every rule for $\neg p$. However, rules 1 and 3 together are superior to 2 and 4 and in BDL defeat 2 and 4. Specifically, for every rule for $\neg p$ there is a superior rule for p and so, when team defeat is adopted, p is derivable. It is not entirely clear that team defeat is always reasonable, since for every argument in support of p there is also an equally good argument against it. NDL and ADL do not possess it as a feature (in NDL and ADL, every rule fends for itself). However, we should mention that team defeat bears a marked similarity to the notion of *admissible set* in Dung's argumentation frameworks [Dun95] (and this similarity suggests that team defeat is reasonable). A set A of arguments is admissible *wrt* to another set B if for every argument $a \in A$ attacked by $b \in B$, there is an $a' \in A$ that attacks b. In the terminology of defeasible logic, the admissible set A constitutes a team of rules.

Team defeat should not be confused with Pollock's *collective defeat*, in which competing rules defeat each other (and which thus can be used to block ambiguity)[Pol87]. With collective defeat, sets of rules defeat each other. With team defeat, gangs of rules defeat some other gang.

3.3.2 THE LOGIC BDLA

The ambiguity propagating logic BDLA is created by adding to BDL a further tag Σ to indicate when a literal is supported (though possibly still defeated). It is this tag which allows the construction of something akin to the "zombie paths" discussed by Makinson and Schlecta [MS91]. The

modifications to the proof system are shown below. Conditions 3.1.3.1 and 4.1.3.1 are changed, and conditions 5 and 6 are added (leading to a considerably more involved proof-system).

Definition 3.18. ***(Proof in*** $BDLA$***)*** *τ is a* defeasible proof *in $BDLA$ iff τ is an argument tree for D, and for each node n of τ, one of the following obtains.*

1. $label(n) = +\Delta p$ *and there is an* $r \in R_s[p]$ *and* $body(r)$ *strictly succeeds at* n.
2. $lable(n) = -\Delta p$ *and for all* $r \in R_s[p]$, $body(r)$ *strictly fails at* n.
3. $label(n) = +\delta p$ *and either*

   3.1 $\{p\}$ *strictly succeeds at* n, *or*

   3.1.1 *there exists an* $r \in R_{sd}[p]$ *such that* $body(r)$ *succeeds at* n, *and*
   3.1.2 $\{\neg p\}$ *strictly fails at* n, *and*
   3.1.3 *for all rules* $s \in R[\neg p]$, *either*

   3.1.3.1 $body(s)$ *is not supported at* n, *or*
   3.1.3.2 *there exists a* $t \in R[p]$ *such that*

   3.1.3.2.1 $body(t)$ *succeeds at* n, *and* $s \prec t$
4. $label(n) = -\delta p$ *and*

   4.1 $\{p\}$ *strictly fails at* n, *and either*

   4.1.1 *for all* $r \in R_{sd}[p]$ $body(r)$ *fails at* n, *or*
   4.1.2 $\{\neg p\}$ *strictly succeeds at* n, *or*
   4.1.3 *there exists a* $s \in R[\neg p]$ *such that*

   4.1.3.1 $body(s)$ *is supported at* n, *and*
   4.1.3.2 *for all* $t \in R[p]$, *either*

   4.1.3.1 $body(t)$ *fails at* n, *or* $s \not\prec t$
5. $label(n) = +\Sigma p$ *and either*

   5.1 $\{p\}$ *strictly succeeds at* n, *or*
   5.2 *there exists an* $r \in R_{sd}[p]$ *such that*

   5.2.1 r *is supported at* n, *and*
   5.2.2 *for all* $s \in R[\neg p]$, *either*

   5.2.2.1 $body(s)$ *fails at* n, *or* $r \not\prec s$.
6. $label(n) = -\Sigma p$ *and*

   1.1 $\{p\}$ *strictly fails at* n, *and*
   6.2 *for all* $r \in R_{sd}[p]$, *either*

6.2.1 *r is not supported at n, or*
6.2.2 *there exists a $s \in R[\neg p]$ such that*

6.2.2.1 *$body(s)$ succeeds at n, and $r \prec s$.*

The above criteria for $+\Sigma p$ and $-\Sigma p$ first appeared in [ABGM00b] in the form of a metaprogram—the proof system was cached out as a logic program interpreted under the Kunen semantics [Kun87]. A more formal presentation of the proof system is found in [ABG$^+$00]. Again, we note that the proof systems do not involve trees (as present here); proofs are sequences of tagged literals.

Without priorities, conditions 5 and 6 simplify to

5. $label(n) = +\Sigma p$ and there exists a rule $r \in R_{sd}[p]$ such that $body(r)$ is supported at n.
6. $label(n) - \Sigma p$ and for all rules $r \in R_{sd}[p]$, $body(r)$ is not supported at n.

This simplified presentation can be found in, for instance, [GMAB04]). However, the simplification shows that in unprioritized theories Σ allows the drawing of some very odd conclusions. Consider the unprioritized theory

Example 3.19.

(1) $\{\} \rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{\neg p\} \rightarrow r$
(4) $\{\} \Rightarrow \neg r$

The atom p is a fact of the theory and so defeats the presumption for $\neg p$. The only rule for r fails and so we can derive $-\delta r$. However, according to the conditions for $+\Sigma$ and $-\Sigma$, $\neg p$ and thus r are still supported (even though defeated), and this is sufficient to defeat $\neg r$. Obviously, this is not correct. The difficulty is that, in the absence of priorities, the definition of $+\Sigma$ does not distinguish between strict and defeasible rules. In the present example, it should be the case that since r_1 is strict and r_2 defeasible, r_2 is inferior to r_1. If that's the case, then one can conclude $-\Sigma\neg p$ and $-\Sigma r$ and hence $+\delta\neg r$. Unfortunately, a blanket preference of strict rules to defeasible rules cannot be endorsed. In the below modified example, since r_1 is supported by a literal only defeasibly derivable, there is no reason to prefer r_1 to r_2. And so the correct conclusion in that case *is* $-\delta\neg r$.

Example 3.20.

(1) $\{q\} \rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{\neg p\} \rightarrow r$
(4) $\{\} \Rightarrow \neg r$
(5) $\{\} \Rightarrow q$

The point to take away from this discussion appears to be that, as $+\Sigma$ is defined in BDLA, the priority relation must be used in some cases to adequately handle conflicts between strict and defeasible rules, and it must be used on a case-by-case basis. It is true that priorities can be compiled away in a sense—as was shown in [ABGM00c] [GMAB04], a prioritized defeasible theory can be translated into an equivalent un-prioritized theory. However, priorities are still needed in the original theory and so the problem is not resolved.

Note that the method of achieving ambiguity propagation in BDLA is different than that used in ADL, and in particular it yields a proof system that is considerably longer than that of BDL. One could modify BDL in precisely the same manner as NDL was modified to produce ambiguity propagating behavior. Indeed, if one requires that a conflicting rule s be superior to r in order to defeat r, then BDL would have precisely the same results in Example 3.13 as ADL. In Example 3.20, $+\delta q$ would be derivable. Nothing about the other literals can be concluded. In the absence of extended conflict sets, this is what is desired.

3.4 Relative Provability

In this section we briefly compare the proof-systems. Several propositions and two tables are presented which indicate relative provability in the logics. The examples supporting these results are provided in Appendix 3. In the tables, NDL_{ext} and ADL_{ext} are used to indicate that conflict sets are closed in those logics; the absence of the subscripts means that conflict sets are taken as minimal. Making this distinction is necessary (because BDL and BDLA do not have extended conflict sets).

For the sake of comparisons, we only concern ourselves with acyclic unprioritized theories. By acyclic, we simply mean that the dependency graph of each literal in the defeasible theory contains no cycles. The restriction is made so that failure-by-looping does not apply (recall that failure-by-looping is absent from BDL and BDLA). We also assume that $+\Sigma$, $-\Sigma$, $+\Delta$ and $-\Delta$ are defined for NDL and ADL. This is safe, since the notion of defeasible proof in ADL and NDL does not depend on that of strict proof, and the notion of strict proof is the same regardless of the proof system considered. Furthermore, without priorities, the definition of support reduces to the simplified criteria from the previous section; it is completely self contained. In any of the logics L above, $Cl_L(D)$) denotes the set of tagged literals that appear as root of a valid proof tree over D in L.

Propositions. *Let D be a finite, acyclic defeasible theory with an empty priority relation and $L1$ and $L2$ one of NDL, ADL, BDL, or BDLA (for BDL and BDLA, assume that team defeat is not used). The following propositions hold.*

P1. $+\Delta p \in Cl_{L1}(D)$ iff $+\Delta p \in Cl_{L2}(D)$.
P2. $-\Delta p \in Cl_{L1}(D)$ iff $-\Delta p \in Cl_{L2}(D)$.
P3. $+\Sigma p \in Cl_{L1}(D)$ iff $+\Sigma p \in Cl_{L2}(D)$.
P4. $-\Sigma p \in Cl_{L1}(D)$ iff $-\Sigma p \in Cl_{L2}(D)$.
P5. If $+\Delta p \in Cl_{L1}(D)$, then $+\delta p \in Cl_{L1}(D)$.

If $+\delta p \in Cl_{L1}(D)$, then $+\Sigma p \in Cl_{L1}(D)$.
P6. If $-\Sigma p \in Cl_{L1}(D)$, then $\{-\Delta p, -\delta p\} \subseteq Cl_{L1}(D)$.
P7. If $+\delta p \in Cl_{BDLA}(D)$, then $+\delta p \in Cl_{NDL}(D)$ and $+\delta p \in Cl_{ADL}(D)$.
P8. If $-\delta p \in Cl_{NDL}(D)$ or $-\delta p \in Cl_{ADL}(D)$, then $-\delta p \in Cl_{BDLA}(D)$.
P9. If $+\delta p \in Cl_{BDLA}(D)$, then $+\delta p \in Cl_{BDL}(D)$. [ABG$^+$00]
P10. If $-\delta p \in Cl_{BDL}(D)$, then $-\delta p \in Cl_{BDLA}(D)$. [ABG$^+$00]

P1, P2, P3, P4, P5, and P6 are all trivial to prove. P1 and P2 hold because the definition of $+\Delta$ and $-\Delta$ are the same for all of the logics. The same holds for $+\Sigma$ and $-\Sigma$ and P3 and P4. P5 can be proven by induction on the depth of strict proof trees; from each strict tree with head $+\Delta p$, a new tree with head $+\delta p$ can be created. Intuitively, P6 is correct, since it is impossible to either definitely or defeasibly prove a literal that isn't even supported. P9 and P10, which relate

provability and refutability in BDL to BDLA are stated in [ABG$^+$00] and [ABGM00b]. Proofs for P7 and P8, which relate BDLA to NDL and ADL appear in Appendix A.

Tables 3.1 and 3.2 present further results. From the tables, it can be seen that no defeasible logic subsumes another.

There are a few additional comments that can be made regarding the logics. The differences between BDL and NDL are due to many factors (conflict sets,team defeat, the handling of strict rules with priorities, failure by looping, etc). Under restrictions, however, the consequences of the logics do coincide.

Proposition 3.21. *Let D be an acyclic defeasible theory without strict rules and no team defeat. Assume conflict sets are minimal. Then for any literal p, $D \mid\sim_{BDL} p$ iff $D \mid\sim_{NDL} p$ and $D \sim\!\!\mid_{BDL} p$ iff $D \sim\!\!\mid_{NDL} p$.*

Proof. In the absence of strict rules, the tags $+\Delta$ and $-\Delta$ become irrelevant. Since D is acyclic, failure-by-looping does not apply. By examination of the proof systems of BDL and NDL, it can be seen that the two systems coincide. □

No general correspondence exists between $BDLA$ and ADL. In the theory

Example 3.22.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$

$BDLA$ concludes both $-\delta p$ and $-\delta\neg p$, whereas ADL concludes nothing at all. This is actually somewhat interesting, as it's possible for a literal in $BDLA$ to be refuted and yet still be used in a proof to defeat other literals. This can happen in no other defeasible logic.

3.5 Historical Developments

Of the logics discussed in this chapter, BDL was the first to appear in print (1993) [Bil93]. Most of the features of logics based upon it are present: the use of tagged literals to distinguish strict and defeasible provability/refutability, the potential defeat of strict rules by defeasible ones, a linear

Y

X	BDL	$BDLA$	NDL	NDL_{ext}	ADL	ADL_{ext}
BDL	YES	NO[1]	NO[1]	NO[1]	NO[1]	NO[1]
$BDLA$	YES[d]	YES	YES[b]	NO[1]	YES[b]	NO[1]
NDL	NO[1]	NO[1]	YES	NO[1]	NO[7]	NO[1]
NDL_{ext}	NO[4]	NO[1]	NO[1]	YES	NO[1]	NO[1]
ADL	NO[1]	NO[1]	YES[a]	NO[1]	YES	NO[1]
ADL_{ext}	NO[4]	NO[4]	NO[4]	YES[a]	NO[4]	YES

1: Ex A1. 2: Ex A2. 4: Ex A4. 5: Ex A5, 6: Ex A6. 7: Ex A7.
Examples A1–A7 are given in Appendix 3.
a: Proposition 3.15
b: Prop. P7
c: Prop. P8
d: Prop. P9
e: Prop. P10

Table 3.1: For logics X and Y, if $+\delta p \in Cl_X$ then $+\delta p \in Cl_Y$.

Y

X	BDL	$BDLA$	NDL	NDL_{ext}	ADL	ADL_{ext}
BDL	YES	YES[e]	NO[1]	NO[1]	NO[1]	NO[1]
$BDLA$	NO[1]	YES	NO[1]	NO[1]	NO[1]	NO[1]
NDL	NO[1]	YES[c]	YES	NO[1]	NO[1]	NO[1]
NDL_{ext}	NO[1]	NO[1]	NO[1]	YES	NO[1]	NO[1]
ADL	NO[5]	YES[c]	YES[a]	No[6]	YES	NO[1]
ADL_{ext}	NO[4]	NO[4]	NO[4]	YES[a]	NO[2]	YES

1: Ex A1. 2: Ex A2. 4: Ex A4. 5: Ex A5, 6: Ex A6. 7: Ex A7.
Examples A1–A7 are given in Appendix 3.
a: Proposition 3.15
b: Prop. P7
c: Prop. P8
d: Prop. P9
e: Prop. P10

Table 3.2: For logics X and Y, if $-\delta p \in Cl_X$ then $-\delta p \in Cl_Y$.

proof system, minimal conflict sets, and team defeat. The idea of *strong negation* appears there as well—the specifications for $-\Delta$ and $-\delta$ are *in a sense* the negation of those for $+\Delta$ and $+\delta$. *E.g.*, one may negate the specification for $+\delta p$ and replace each tag with its complement to get the specification for $-\delta p$.

BDL is itself a modification of an earlier logic appearing in [NBC89] and more prominently in [BCN90]. BDL generalizes this logic to theories involving variables and function symbols (innovations we have chosen to ignore in the present chapter). The only other difference between the two logics appears not in the proof system but in restrictions on the precedence relation. In the earlier logic, strict rules are superior to all defeasible rules and no strict rule is superior to any other strict rule (this is how it is in ADL and NDL). BDL allows a defeasible rule to have priority over a strict rule. In his analysis of BDLA without team defeat, Brewka shows that when the precedence relation is restricted to defeasible rules (no defeasible rule is superior to a strict one), then BDLA is sound (but not complete) *wrt* his prioritized well-founded semantics [Bre01] [Bre96]. He does not believe that the precedence relation should be defined over strict rules.

In [ABM98] it is shown that the superiority relation can be removed from BDL without limiting the expressiveness of the logic—theories with a nonempty superiority relation can be translated into an equivalent one with an empty relation. The transformations are *modular* in the sense that adding more rules does not require a complete retranslation of the original theory. It is noted that the modularity is surely due to the use of minimal conflict sets (this can be readily seen by examining the transformation scheme), and this desirable property is lost if conflict sets are closed under strict rules. [ABGM00c] shows that defeaters and superiority relations can be removed in linear time and that the resulting theory is at most 12 times the size of the original in terms of number of symbols.

The last result is significant, since it is shown in [Mah01] (also [MRA$^+$01]) that the conclusions under BDL of unprioritized theories without defeaters can be computed in linear time *wrt* to the size of the theory. The algorithm is similar to that found in [DG84], which showed that the least Herbrand model of definite logic programs can be computed in linear time. It should be noted again that this result holds if the precedence relation is acyclic and conflict sets are minimal.

Ambiguity propagation appeared in defeasible logic around the year 2000 [ABGM00b]. Up to that point, all defeasible logics were ambiguity blocking. In [ABGM00b] the propagating logics are presented as systems embedded in logic programs. A proof system later appeared in [ABG$^+$00] and [AB01].

Historically, defeasible logic has been defined in terms of proof systems. However, several alternative semantics for BDL and related logics have been provided over the years. A Dung-like argumentation semantics is provided in [GMAB00] (an extended version appears as [GMAB04]). A model theory for BDL is discussed in [Mah02]. Fixpoint semantics for BDL is provided in [MG99]; the proof system is sound and complete with respect to this semantics. The same paper presents a well-founded semantics for defeasible logic. However, this semantics does not agree in general with the WFS or stable model semantics for logic programs under the most natural translation (the translation scheme discussed in Chapter 5), even for theories without priorities. Particularly, like the proof systems upon which it is based, this semantics does not (in our opinion) handle indirect conflicts correctly.

From a practical standpoint, some of the more interesting work in defeasible logic has been in studying its relations to other formalisms. [AB01] gives a way of translating defeasible logic into default logic such that every defeasible consequence of the defeasible theory is in every extension of the default theory. The paper does not appear to address refutations (it is not proven whether a literal refuted in defeasible logic appears in no default extension). Embedding (translating) defeasible logic into logic programs appears in [MG99] [ABGM00b] and most recently in [ABGM06]. In [ABGM06], it is shown that under the translation, the conclusions of the defeasible theory correspond to the intersection of stable models. However, this results holds only for what the authors call *decisive* theories—theories in which every literal is either provable or refutable. Without decisiveness, the correspondence holds only in one direction: every literal provable in defeasible logic is in the intersection of stable models. In the general case, the results of defeasible logic correspond to consequences of the logic program under the Kunen semantics [Kun87]. One might consider this to provide an indirect semantics for defeasible theories.

Regarding the northern variants of defeasible logic, NDL and ADL incorporate extended conflict sets (to handle indirect conflicts) and loop-checking, both of which are absent from the logics based on BDL. Nute cites [Sch94] as the motivation for extended conflict sets [Nut94][Nut97]. NDL and ADL also do not incorporate team defeat. In comparison to BDL and its offshoots, there historically has been little work done to provide a semantics for NDL. Donnelly in [Don99] provides a fixpoint semantics that is not too dissimilar to stable models or the extensions of default logic. It is indirectly skeptical. Consequences are defined as the intersection of the stable sets (which Donnelly calls T-extensions). Floating conclusions are possible and hence NDL's proof system is not complete *wrt* to the semantics. The proof system is sound, however.

3.6 Problematic Examples

The defeasible logics discussed so far all yield unintuitive results for a large class of theories, and the results in these cases are sufficiently troubling to warrant altering the logics. As was discussed, BDL and BDLA cannot detect indirect conflicts nor can they dismiss circular arguments. ADL and NDL do not allow reinstatement, a principle widely endorsed. Consider the below theory.

Example 3.23.

(1) $\{\} \rightarrow married(chris)$
(2) $\{\} \Rightarrow \neg married(chris)$
(3) $\{\} \Rightarrow husband(chris)$
(4) $\{husband(chris)\} \rightarrow married(chris)$

With conflict sets closed under strict rules, $\neg married(chris)$ conflicts with $husband(chris)$. In the theory, $married(chris)$ is a fact and consequently defeats the presumption $\neg married(chris)$. The body of rule 2 succeeds, however, and this is sufficient to prevent deriving $husband(chris)$ in NDL and ADL. The conclusions of this theory according to the ambiguity blocking logic NDL are $+\delta married(chris)$, $-\delta husband(chris)$ and $-\delta\neg married(chris)$. For the propagating logic ADL, $husband(chris)$ is undetermined. Neither of these results is particularly acceptable, for the only rule conflicting with r_3 and preventing deriving $husband(chris)$ is r_2, and r_2 is defeated by r_1.

Note that if the contrapositive of rule 4 (*i.e.* , $5.\neg married(chris) \rightarrow \neg husband(chris)$) is added to the theory as an extra rule and conflict sets are considered minimal, then the intuitively correct result is obtainable in both NDL and ADL. Rule r_2 is defeated by r_1. r_5 fails and thus no successful rule conflicts with r_3. This allows the conclusion $+\delta husband(chris)$ to be drawn. As will be discussed in Chapter 6, for unprioritized theories, adding the transposition of strict rules while keeping conflict sets minimal will allow the intuitively correct results to be drawn from a theory. Doing this can be done more efficiently than explicitly keeping track of extended conflict sets.

Chapter 6 defines new logics which are designed to overcome the problems that NDL and ADL suffer from. However, we first present fixpoint semantics for NDL and ADL (Chapter 4) and their relation to logic programming (chapter 5). Semantics for the new logics will be based upon this material.

CHAPTER 4

WELL-FOUNDED SEMANTICS FOR ADL AND NDL

Sam Donnelly in [Don99] provided a fixpoint semantics for the logic NDL and showed that the proof system for NDL was sound but not complete *wrt* to this semantics, even for finite theories. The result is not surprising. In defeasible logic, as in the WFS, in order for a literal to be provable (or "well-founded"), there must be some rule for it whose body is also provable. Donnelly's semantics is based on the intersections of extensions and so allows floating conclusions. It is unlikely that a constructive proof system would be complete with respect to such a semantics.

In this chapter, inspired by the well-founded semantics for logic programs, we provide fixpoint semantics for both NDL and ADL. It will be shown that the proof systems for NDL and ADL are sound relative to their counterpart semantics. We define the *finite grounded components* of a theory's well-founded model, and we show that completeness holds *wrt* these finite grounded components. Every literal that is well-founded or unfounded in a finite grounded component has a corresponding proof tree in the proof system. For finite theories, the well-founded model is itself a finite grounded component, and so for finite theories the proof systems are complete *wrt* the semantics. *Locally finite* theories are also defined; completeness is shown for this more general class as well.

Later sections of the chapter show that NDL and ADL both fail Cut and Cautious Monotony. However, if the precedence relation over rules is transitive, then the properties are satisfied. It is also shown that well–founded literals can be deleted from the bodies of rules without affecting the consequences of a theory, and rules with unfounded bodies can be deleted from the theory. Two antimonotone operators are also defined. One, which we call α, propagates ambiguity, while the other, β, blocks it. It is shown that β can be used to compute the consequences of theories

according to NDL. In Chapter 5, it is shown that under some restrictions, α can be used to compute consequences according to ADL.

4.1 Unfounded Sets and the Well-Founded Model

Let D be a defeasible theory. An *interpretation* $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ is an ordered pair of disjoint subsets of Lit_D. To pick out a particular interpretation we will sometimes use subscripts, as in $\mathcal{I}_n = \langle \mathcal{T}_n, \mathcal{U}_n \rangle$. An interpretation can be viewed as partitioning Lit_D into three sets: The well-founded, the unfounded, and the undetermined. Interpretations can be partially ordered as follows: $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$ if and only if both $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{U}_1 \subseteq \mathcal{U}_2$.

In defeasible logic, unfounded sets are collections of literals for which no external support exists. The only way to prove an unfounded literal is to use literals that are themselves unfounded. The basic idea is that a set of literals S is unfounded if for every $p \in S$, every strict rule for p has a subgoal that is already known to be unfounded (that is, appears in $\mathcal{U}$) or else is in S itself. The latter prevents circular reasoning from justifying p. The same holds for defeasible rules, with the addition that a defeasible rule can be defeated by sets of superior rules. We formalize this idea below.

Definition 4.1 (Unfounded Sets in ADL)**.** *A set S is* unfounded *in ADL with respect to D and an interpretation $\mathcal{I}$ iff for all literals $p \in S$:*

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U} \cup S) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

- *(2.1) $body(r) \cap (\mathcal{U} \cup S) \neq \varnothing$, or*
- *(2.2) there is a $c \in C[p]$ such that for each $q \in c - \{p\}$ there is a rule $s \in R[q]$ such that*
  - *(2.2.1) $body(s) \subseteq \mathcal{T}$ and*
  - *(2.2.2) $r \prec s$ or s is strict.*

Definition 4.2 (Unfounded Sets in NDL)**.** *The definition of unfounded set in NDL is exactly the same as for ADL, save that condition 2.2.2 is replaced with the following requirement: $s \not\prec r$.*

Example 4.3.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{p\} \Rightarrow r$
(4) $\{q\} \Rightarrow q$

In the above example, $\{p, \neg p\}$, $\{q\}$, and $\{p, \neg p, q, r\}$ are all unfounded under NDL with respect to $\langle \varnothing, \varnothing \rangle$. However, only $\{q\}$ is unfounded under ADL. Since neither rule 1 nor 2 is preferred to the other, $\{p, \neg p\}$ is not unfounded under ADL, and since p is not unfounded, neither is r.

Lemma 4.4 (Unfounded sets are closed under union)**.** *If S is a set of unfounded sets* wrt *to defeasible theory D and interpretation $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$, then $\bigcup_{i=0}^{\infty}\{S_i |\ S_i \in S\}$ is unfounded* wrt *to D and $\mathcal{I}$.*

Proof. Let $V = \bigcup_{i=0}^{\infty}\{S_i |\ S_i \in S\}$ and suppose $p \in V$. Let $S_j \in S$ such that $p \in S_j$. If $r \in R_s[p]$, then since S is unfounded, $body(r) \cap (S_j \cup \mathcal{U}) \neq \varnothing$. But since $S \subseteq V$, $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$. Similarly, if $r \in R_d[p]$, then either $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$ or there is a $c \in C[p]$ such that for all $u \in c - \{p\}$, there is some rule s such that $body(s) \subseteq \mathcal{T}$ and $r \prec s$ or else s is strict (in NDL, $s \not\prec r$). Generalizing on p, it follows that V is unfounded with respect to D and $\mathcal{I}$. □

Since unfounded sets are closed under union, then given a defeasible theory D and interpretation $\mathcal{I}$, there exists a greatest unfounded set *wrt* D and $\mathcal{I}$. We define this as an operator.

Definition 4.5. $U_D(\mathcal{I}) = \bigcup\{S |\ S$ *is an unfounded set* wrt *to D and $\mathcal{I}$*$\}$.

The operator U_D can be viewed as producing the set of literals that are unfounded with respect to an interpretation. The immediate consequence operator T_D, defined below for defeasible theories, generates the set of literals that are well-founded with respect to that interpretation. Lemma 4.9 shows that both U_D and T_D are monotone.

Definition 4.6 (Witness of Provability)**.** *Let D be a defeasible theory and $\mathcal{I}$ an interpretation. A rule $r \in R_D$ is a* witness of provability *for p* wrt *D and $\mathcal{I}$ if one of the below conditions applies.*

(1) $r \in R_s[p]$ *and* $body(r) \subseteq \mathcal{T}$.

(2) $r \in R_d[p]$ *and* $body(r) \subseteq \mathcal{T}$*, and for each conflict set* $c \in C[p]$*, there exists a* $q \in c - \{p\}$ *such that for all* $s \in R[q]$*,* $s \prec r$ *or* $body(s) \cap \mathcal{U} \neq \varnothing$*.*

Definition 4.7 (Immediate Consequence for ADL and NDL)**.** *Let* D *be a defeasible theory and* $\mathcal{I}$ *an interpretation. The immediate consequences of* D wrt $\mathcal{I}$*, written* $T_D(\mathcal{I})$ *is the set*

$$T_D(\mathcal{I}) = \{p| \textit{ there exists a witness of provability for } p \text{ wrt } D \textit{ and } \mathcal{I}\}.$$

Definition 4.8 (Supported Rule)**.** *If* r *is a rule and* $\mathcal{I}$ *an interpretation, then* r *is supported in* $\mathcal{I}$ iff $body(r) \subseteq \mathcal{T}$*.*

If r has an empty body, then because it is supported in every interpretation, we will sometimes speak loosely, saying that r is supported without explicitly referring to any interpretation.

Lemma 4.9 (T and U are monotone)**.** *Let* D *be a defeasible theory and* $\mathcal{I}_1$ *and* $\mathcal{I}_2$ *interpretations. If* $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$*, then* $T_D(\mathcal{I}_1) \subseteq T_D(\mathcal{I}_2)$*, and* $U_D(\mathcal{I}_1) \subseteq U_D(\mathcal{I}_2)$*.*

Proof. Suppose $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$ and let $p \in T_D(\mathcal{I}_1)$. Then either there exists an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_1$, or there exists a rule $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_1$ and for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R[q]$, $s \prec r$ or $body(s) \cap \mathcal{U}_1 \neq \varnothing$. Since $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$, $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{U}_1 \subseteq \mathcal{U}_2$. By substitution, we have $r \in R_s[p]$, and $body(r) \subseteq \mathcal{T}_2$, or $r \in R_d[p]$ and $body(r) \subseteq \mathcal{T}_2$ and for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R[q]$, $s \prec r$ or $body(s) \cap \mathcal{U}_2 \neq \varnothing$. It can be seen that $p \in T_D(\mathcal{I}_2)$. Generalizing on p, $T_D(\mathcal{I}_1) \subseteq T_D(\mathcal{I}_2)$.

Suppose now that $p \in U_D(\mathcal{I}_1)$. Since $U_D(\mathcal{I}_1)$ is the greatest unfounded set *wrt* D and $\mathcal{I}_1$,

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U}_1 \cup U_D(\mathcal{I}_1)) \neq \varnothing$.
(2) For every $r \in R_d[p]$,
  (2.1) $body(r) \cap (\mathcal{U}_1 \cup U_D(\mathcal{I}_1)) \neq \varnothing$, or
  (2.2) There is a $c \in C[p]$ such that for each $q \in c - \{p\}$ there is a rule $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_1$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$).

Substituting $\mathcal{T}_2$ for $\mathcal{T}_1$ and $\mathcal{U}_2$ for $\mathcal{U}_1$ and generalizing on p, we see that $U_D(\mathcal{I}_1)$ is unfounded *wrt* D and $\mathcal{I}_2$. So by the definition of U_D, $U_D(\mathcal{I}_1) \subseteq U_D(\mathcal{I}_2)$. □

4.1.1 THE WELL-FOUNDED MODEL

Given a theory D and an interpretation $\mathcal{I}$, both $T_D(\mathcal{I})$ and $U_D(\mathcal{I})$ produce sets of literals. The operator $W_D(\mathcal{I})$ combines them to form a new interpretation.

Definition 4.10 (Well-Founded Operator)**.** $W_D(\mathcal{I}) = \langle T_D(\mathcal{I}), U_D(\mathcal{I})\rangle$

W_D defines a sequence of interpretations $(\mathcal{I}_{D,0}, \mathcal{I}_{D,1}, \ldots)$:

$\mathcal{I}_{D,0} = \langle \varnothing, \varnothing\rangle$
$\mathcal{I}_{D,\alpha+1} = W_D(\mathcal{I}_{D,\alpha})$ (for successor ordinals $\alpha + 1$)
$\mathcal{I}_{D,\alpha} = lub(\{\mathcal{I}_{D,\beta} | \beta < \alpha\})$ (where α is a limit ordinal).

Lemma 4.11 ($\mathcal{I}$ is monotone nondecreasing)**.** *Let D be a defeasible theory. For any $\alpha \geq 0$, $\mathcal{I}_{D,\alpha} \sqsubseteq \mathcal{I}_{D,\alpha+1}$.*

Proof. This follows immediately from the definition of the sequence $(\mathcal{I}_D)$ and the monotonicity of T_D and U_D. □

We have insisted that $\mathcal{T}$ and $\mathcal{U}$ be disjoint in a valid interpretation. However, eliminating this restriction causes the set of interpretations to become a complete lattice under the $\sqsubseteq$ relation. Since W_D is monotone, then by the Knaster-Tarski Theorem (2.25), least and greatest fixpoints of W_D exist. We take the least fixpoint to be the well–founded model of a defeasible theory.

Definition 4.12 (The Well-Founded Model)**.** *Let D be a defeasible theory. The* well-founded model *of D, written $wfm(D)$, is $lfp(W_D)$.*

We will sometimes write the well-founded model of D as $\mathcal{I}_{D,WF}$ to emphasize that it is an interpretation, and often we will eliminate the theory in the subscript (writing $\mathcal{I}_{WF}$) when the theory is not in doubt.

Definition 4.13. *Let D be a defeasible theory, L one of NDL or ADL, and $\mathcal{I}_{WF} = \langle \mathcal{T}_{WF}, \mathcal{U}_{WF}\rangle$ D's well-founded model under L. Define $D \mid\!\approx_L p$ to mean $p \in \mathcal{T}_{WF}$ under L and $D \approx\!\!\mid_L p$ to mean $p \in \mathcal{U}_{WF}$ under L.*

Importantly, the sequence of interpretations defined using W_D is coherent, in the sense that for any literal p, p cannot both be in $\mathcal{T}$ and $\mathcal{U}$.

Lemma 4.14 (The Sequence $(\mathcal{I})$ is coherent)**.** *Let D be a defeasible theory and $(\mathcal{I}_D)$ the sequence of interpretations defined by W_D by iterating from $\langle \varnothing, \varnothing \rangle$. For any $\lambda \geq 0$, $\mathcal{T}_{D,\lambda} \cap \mathcal{U}_{D,\lambda} = \varnothing$.*

Proof. For $\lambda = 0$, the lemma holds. Suppose it holds for all $\alpha < \lambda$ and let λ be a successor ordinal. Let A be some set of literals such that $A \cap \mathcal{T}_\lambda \neq \varnothing$. Since $A \cap \mathcal{T}_\lambda \neq \varnothing$, there must be some $\kappa \leq \lambda$ such that $\mathcal{I}_\kappa$ is the earliest in the sequence $(\mathcal{I}_0, \mathcal{I}_1, \ldots)$ for which $\mathcal{T}_\kappa \cap A \neq \varnothing$. For all $\iota < \kappa$, $\mathcal{T}_\iota \cap A = \varnothing$.

Let p be a literal such that $p \in A$ and $p \in \mathcal{T}_\kappa$. Then either (Immediate Consequence) definition 4.8.1 or 4.8.2 holds. Suppose it's 4.8.1. Then there is an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_\iota$ for some $\iota < \kappa$. By choice of A, $A \cap \mathcal{T}_\iota = \varnothing$. So, $body(r) \cap A = \varnothing$. By Lemma 4.11 above, since $body(r) \subseteq \mathcal{T}_\iota$, $body(r) \subseteq \mathcal{T}_{\lambda-1}$. By the inductive hypothesis, $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$. Since $p \in A$ and there exists an $r \in R_s[p]$ such that $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$, A fails the definition of unfounded set with respect to D and $\mathcal{I}_\lambda$.

Suppose it's 4.8.2. Then there is an $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_\iota$ for some $\iota < \kappa$ and for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R[q]$, $s \prec r$ or $body(s) \cap \mathcal{U}_\iota \neq \varnothing$. By choice of A, $A \cap \mathcal{T}_\iota = \varnothing$. So, $body(r) \cap A = \varnothing$. By Lemma 4.11 above, since $body(r) \subseteq \mathcal{T}_\iota$, $body(r) \subseteq \mathcal{T}_{\lambda-1}$. Since $body(r) \subseteq \mathcal{T}_{\lambda-1}$, then by the inductive hypothesis $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$. Also by Lemma 4.11, if $body(s) \cap \mathcal{U}_\iota \neq \varnothing$, then $body(s) \cap \mathcal{U}_{\lambda-1} \neq \varnothing$. If this is so, then by the inductive hypothesis, $body(s) \nsubseteq \mathcal{T}_{\lambda-1}$. Generalizing on s, there exists a $p \in A$ and a rule $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_{\lambda-1}$ and for all conflict sets $c \in C[p]$ there is a q for which for all rules $s \in R[q]$, $s \prec r$ or else $body(s) \nsubseteq \mathcal{T}_{\lambda-1}$. A again violates the definition of unfounded set with respect to D and $\mathcal{I}_{\lambda-1}$.

Generalizing on A, no set A that intersects $\mathcal{T}_\lambda$ is unfounded with respect to D and $\mathcal{I}_{\lambda-1}$. In other words, if a set A is unfounded *wrt* $\mathcal{I}_{\lambda-1}$, then it does not intersect $\mathcal{T}_\lambda$. In particular, $\mathcal{U}_\lambda$—the greatest unfounded set with respect to D and $\mathcal{I}_{\lambda-1}$—does not intersect $\mathcal{T}_\lambda$. We conclude that for any $p \in \mathcal{T}_\lambda$, $p \notin \mathcal{U}_\lambda$.

If λ is a limit ordinal, then if $\mathcal{T}_\lambda \cap \mathcal{U}_\lambda \neq \varnothing$, for some p we have $p \in \mathcal{T}_\lambda$ and $p \in \mathcal{U}_\lambda$. By definition of $\mathcal{I}_\lambda$, there must be a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_\kappa$ and $p \in \mathcal{U}_\kappa$. This contradicts the assumption that the hypothesis holds for all $\alpha < \lambda$, and so for all $p \in \mathcal{T}_\lambda$ we have $p \notin \mathcal{U}_\lambda$. □

4.1.2 EXAMPLES

In general, the proof systems for ADL and NDL are incomplete *wrt* the semantics just defined.

Example 4.15. *Let D be the following infinite defeasible theory (assume no precedence relation):*

$\{\} \Rightarrow p$
$\{q_0\} \Rightarrow \neg p$
$\{q_{n+1}\} \Rightarrow q_n$ (for each $n \in \mathbb{N}$)

It is clear that neither the proof system for NDL nor the one for ADL is capable of showing $D \mid\!\sim p$ (the refutation of $\neg p$ involves an infinite branch, and proof trees must be finite). However, from the standpoint of the two fixpoint semantics as we have defined them, the set $\{\neg p, q_0, q_1, q_2, \ldots\}$ is unfounded with *wrt* to D and $\mathcal{I}_0$. That is, $\{\neg p, q_0, q_1, q_2, \ldots\} \subseteq \mathcal{U}_1$. As this is so, the rule $\{\} \Rightarrow p$ is allowed to fire, and so $p \in \mathcal{T}_2$. The well-founded model of this example under both semantics is $\langle\{p\}, \{\neg p, q_0, q_1, q_2, \ldots\}\rangle$.

Example 4.16.

(1) $\{\} \rightarrow bird$
(2) $\{\} \rightarrow penguin$
(3) $\{bird\} \Rightarrow flies$
(4) $\{penguin\} \Rightarrow \neg flies$

With no preferences among the rules, the well-founded model according to NDL is $\langle\{bird, penguin\}, \{flies, \neg flies\}\rangle$. For ADL it's $\langle\{bird, penguin\}, \{\}\rangle$. Both of these results agree with the proof systems defined in Chapter 3. If the preference $3 \prec 4$ is added, both ADL and NDL conclude $\langle\{bird, penguin, \neg flies\}, \{flies\}\rangle$. The sequence of interpretations generating this result is

$\mathcal{I}_0 = \langle\{\}, \{\}\rangle$
$\mathcal{I}_1 = \langle\{bird, penguin\}, \{\}\rangle$
$\mathcal{I}_2 = \langle\{bird, penguin, \neg flies\}, \{flies\}\rangle$

Importantly, while W_D is monotone, it is not continuous. In the example below, assume that conflict sets are minimal.

Example 4.17 (W_D is not continuous)**.**

(1) $\{\} \rightarrow p_0$
(2) $\{p_n\} \rightarrow p_{n+1}$ *(for each* $n \in \mathbb{N}$*)*
(3) $\{\neg p_n\} \rightarrow q$ *(for each* $n \in \mathbb{N}$*)*
(4) $\{\} \Rightarrow \neg p_n$ *(for each* $n \in \mathbb{N}$*)*

The set of interpretations produced according to either ADL or NDL is

$\mathcal{I}_0 = \langle\{\}, \{\}\rangle$
(1) $\mathcal{I}_1 = \langle\{p_0\}, \{\neg p_0\}\rangle$
(2) $\mathcal{I}_2 = \langle\{p_0, p_1\}, \{\neg p_0, \neg p_1\}\rangle$
(3) $\mathcal{I}_3 = \langle\{p_0, p_1, p_2\}, \{\neg p_0, \neg p_1, \neg p_2\}\rangle$
(4) *etc.*

In general, for $i > 0$, $\mathcal{I}_i = \langle\{p_0, \ldots p_{i-1}\}, \{\neg p_0, \ldots \neg p_{i-1}\}\rangle$. One sees that

$$\mathcal{I}_\omega = \langle\{p_n | n \in \mathbb{N}\}, \{\neg p_n | n \in \mathbb{N}\}\rangle$$

The literal q is not included in $\mathcal{I}_\omega$. Applying W_D to $\mathcal{I}_\omega$ yields

$$W_D(\mathcal{I}_\omega) = \langle\{p_n | n \in \mathbb{N}\}, \{\neg p_n | n \in \mathbb{N}\} \cup \{q\}\rangle$$

This is the least fixpoint of W_D, and it is not reached until $\mathcal{I}_{\omega+1}$.

4.2 Soundness and (In)Completeness

Though completeness does not hold in general, below we provide a characterization of the circumstances in which well–founded and unfounded literals *do* have corresponding proof trees in the proof systems. In general, literals are provable or refutable using the proof system *iff* they are well–founded or unfounded in some *finite grounded component* of the well-founded model. For finite theories, the well-founded model is itself such a component, and so for those theories the consequences according to proof systems coincide with those according to the semantics.

4.2.1 FINITE GROUNDED COMPONENTS

For the following lemmas, let D be a defeasible theory and $(\mathcal{I}_D)$ the sequence of interpretations produced by iterating W_D from the empty interpretation $\langle \varnothing, \varnothing \rangle$.

Definition 4.18 (Finite Grounded Component)**.** *Finite grounded components are defined inductively.*

(1) $\mathcal{I}_{D,0}$ *is a finite grounded component of* $\mathcal{I}_{D,0}$.
(2) A finite interpretation $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\alpha}$ *is a finite grounded component of* $\mathcal{I}_{D,\alpha}$ *iff for some* $\beta < \alpha$ *there exists a finite grounded component* $\mathcal{Y}$ *of* $\mathcal{I}_{D,\beta}$*, and*
  (2.1) $\mathcal{T}_\mathcal{X} \subseteq T_D(\mathcal{Y})$*, and*
  (2.2) $\mathcal{U}_\mathcal{X}$ *is an unfounded set* wrt D *and* $\mathcal{Y}$.

Observe that $\{\}$ is vacuously an unfounded set *wrt* D and any interpretation $\mathcal{I}$, and also that $\{\} \subseteq T_D(\mathcal{I})$ for any $\mathcal{I}$. And so, for any $\alpha \geq 0$, if $\mathcal{Y}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then if $p \in T_D(\mathcal{Y})$, $\langle \{p\}, \{\} \rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha+1}$. If a finite set S is unfounded *wrt* D and $\mathcal{Y}$, then $\langle \{\}, S \rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha+1}$.

Definition 4.19 (Finite Well-Foundedness/Unfoundedness)**.** *A literal* p *is finitely well-founded (unfounded) at* $\mathcal{I}_{D,\alpha}$ *iff* $p \in \mathcal{T}_\mathcal{X}$ *(*$p \in \mathcal{U}_\mathcal{X}$ *) for some finite grounded component* $\mathcal{X}$ *of* $\mathcal{I}_{D,\alpha}$.

Finite grounded components are stable in the sense that if $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\kappa}$, then $\mathcal{X}$ is a finite grounded component for all later interpretations in the sequence $(\mathcal{I}_D)$.

Lemma 4.20. *For all* $\lambda \geq 0$*, if* $\kappa < \lambda$ *and* $\mathcal{X}$ *is a finite grounded component of* $\mathcal{I}_{D,\kappa}$*, then* $\mathcal{X}$ *is a finite grounded component of* $\mathcal{I}_{D,\lambda}$.

Proof. Since $\kappa < \lambda$, we have $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\kappa} \sqsubseteq \mathcal{I}_{D,\lambda}$. Since $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\kappa}$, there exists a $\iota < \kappa$ and a $\mathcal{Y}$ such that $\mathcal{Y}$ is a finite grounded component of $\mathcal{I}_{D,\iota}$, and

(1) $\mathcal{T}_\mathcal{X} \subseteq T_D(\mathcal{Y})$, and
(2) $\mathcal{U}_\mathcal{X}$ is an unfounded set *wrt* D and $\mathcal{Y}$.

By definition, then, $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\lambda}$. □

Lemma 4.21. *For all $\alpha \geq 0$, if $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then there exists a least successor ordinal $\beta \leq \alpha$ such that $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\beta}$.*

Proof. If α itself is a successor ordinal, then we need show nothing. Suppose α is a limit ordinal. Since $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then by definition of $\mathcal{X}$ there exists a $\lambda < \alpha$ such that $\mathcal{Y}$ is a finite grounded component of $\mathcal{I}_{D,\lambda}$, and

(1) $\mathcal{T}_{\mathcal{X}} \subseteq T_D(\mathcal{T}_{\mathcal{Y}})$, and
(2) $\mathcal{U}_{\mathcal{X}}$ is an unfounded set *wrt* D and $\mathcal{Y}$.

Thus, $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\lambda+1}$. By examination of Definition 4.18, we see that $\mathcal{X}$ is finite grounded component of $\mathcal{I}_{D,\lambda+1}$. □

Like unfounded sets, finite grounded components are closed under (finite) union.

Lemma 4.22. *For all $\alpha \geq 0$, If $\{\mathcal{X}_1, \mathcal{X}_2, \ldots \mathcal{X}_m\}$ is a finite, nonempty set of interpretations such that each $\mathcal{X}_i$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then $\langle \bigcup_{i=1}^m \mathcal{T}_{\mathcal{X}_i}, \bigcup_{i=1}^m \mathcal{U}_{\mathcal{X}_i} \rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$.*

Proof. The proof is by induction on α. $\mathcal{I}_{D,0}$ is empty and so the only nonempty set of finite grounded components is $\{\mathcal{I}_{D,0}\}$, and the least upper bound of this is by definition a finite grounded component of $\mathcal{I}_{D,0}$. Suppose the hypothesis is satisfied by all $\lambda < \alpha$ and consider some set $\{\mathcal{X}_1, \mathcal{X}_2, \ldots \mathcal{X}_m\}$ of finite grounded components of $\mathcal{I}_{D,\alpha}$. From Lemma 4.21, we may assume α is a successor ordinal, and so for each $\mathcal{X}_i$ there exists a $\mathcal{Y}_i$ that is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$. By the inductive hypothesis $\mathcal{Y} = \langle \bigcup_{i=1}^m \mathcal{T}_{\mathcal{Y}_i}, \bigcup_{i=1}^m \mathcal{U}_{\mathcal{Y}_i} \rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$.

Let $\mathcal{X} = \langle \bigcup_{i=1}^m \mathcal{T}_{\mathcal{X}_i}, \bigcup_{i=1}^m \mathcal{U}_{\mathcal{X}_i} \rangle$. For each $\mathcal{Y}_i$, since $\mathcal{Y}_i \sqsubseteq \mathcal{Y}$, we have $T_D(\mathcal{Y}_i) \sqsubseteq T_D(\mathcal{Y})$. However, since by definition of $\mathcal{T}_{\mathcal{X}_i}$, $\mathcal{T}_{\mathcal{X}_i} \subseteq T_D(\mathcal{Y}_i)$, it follows that $\mathcal{T}_{\mathcal{X}_i} \subseteq T_D(\mathcal{Y})$. Generalizing, $\bigcup_{i=1}^m \mathcal{T}_{\mathcal{X}_i} \subseteq T_D(\mathcal{Y})$. Furthermore, since $\mathcal{U}_{\mathcal{X}_i}$ is an unfounded set *wrt* D and $\mathcal{Y}_i$, it follows that $\mathcal{U}_{\mathcal{X}_i}$ is an unfounded set *wrt* D and $\mathcal{Y}$. Generalizing on $\mathcal{X}_i$, since unfounded sets are closed under union, $\bigcup_{i=1}^m \mathcal{U}_{\mathcal{X}_i}$ is an unfounded set *wrt* D and $\mathcal{Y}$.

Thus we have a finite $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\alpha}$, and a finite grounded component $\mathcal{Y}$ of $\mathcal{I}_{D,\alpha-1}$ such that $\mathcal{T}_{\mathcal{X}} \subseteq T_D(\mathcal{Y})$ and $\mathcal{U}_{\mathcal{X}}$ is an unfounded set *wrt* D and $\mathcal{Y}$. $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\alpha}$. □

Lemma 4.23. *If D is finite, then for all $n \geq 0$, $\mathcal{I}_{D,n}$ is a finite grounded component of $\mathcal{I}_{D,n}$*

Proof. The proof is by induction on the length of the sequence $(\mathcal{I}_D)$. Since D is finite, $(\mathcal{I}_D)$ is finite, and so we need only consider integers $n \geq 0$. The claim is true by definition for $n = 0$. Suppose it holds for all $k < n$. By definition of T_D and U_D, $\mathcal{T}_{D,n} = T_D(\mathcal{I}_{D,n-1})$ and $\mathcal{U}_{D,n} = U_D(\mathcal{I}_{D,n-1})$. And so $\mathcal{U}_{D,n}$ is an unfounded set *wrt* D and $\mathcal{I}_{D,n-1}$. However, by hypothesis, $\mathcal{I}_{D,n-1}$ is a finite grounded component of $\mathcal{I}_{D,n-1}$. Thus, by definition of finite grounded components, $\mathcal{I}_{D,n}$ is a finite grounded component of $\mathcal{I}_{D,n}$. □

The below corollary follows immediately from Lemma 4.23 and the fact that $wfm(D)$ is finite.

Corollary 4.24. *If D is finite, then $wfm(D)$ is a finite grounded component of $wfm(D)$.*

Observe that in the above definitions and lemmas, the differences between U_D and T_D as defined for ADL and as defined for NDL are irrelevant. That is, the definitions and lemmas apply to both ADL and NDL. Chapter 6 presents the logics ADL and MDL and their corresponding semantics. The material presented in this section applies to them as well.

4.2.2 Completeness for Finite Grounded Components

Theorem 4.25 (Completeness for Finite Grounded Components)**.** *Let D be a defeasible theory and L one of ADL or NDL. Let $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, ..., be the sequence of interpretations created by iterating W_D from $\langle \varnothing, \varnothing \rangle$. For any $\alpha \geq 0$, If $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then*

(1) if $p \in \mathcal{T}_{\mathcal{X}}$, then $D \mid\sim_L p$, and
(2) if $p \in \mathcal{U}_{\mathcal{X}}$, then $D \sim\!\mid_L p$.

Proof. The proof is by induction on the sequence $(\mathcal{I}_D)$. $\mathcal{I}_0$ is empty and so the hypothesis trivially holds for $\alpha = 0$. Suppose the hypothesis holds for all $\lambda < \alpha$ and suppose α is a successor ordinal. Let $\mathcal{X}$ be a finite grounded component of $\mathcal{I}_{D,\alpha}$.

(Case 1) Suppose $p \in \mathcal{T}_{\mathcal{X}}$. Then there exists a $\mathcal{Y}$ that is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$ and $p \in T_D(\mathcal{Y})$. As such there is some rule $r \in R_{sd}[p]$ such that $body(r) \subseteq \mathcal{T}_{\mathcal{Y}}$. If $r \in R_s[p]$, then by the inductive hypothesis $D \mid\sim_L a$ for each $a \in body(r)$, and so for each there exists a defeasible proof tree with root labeled $+\delta a$. We may append these proofs to a node labeled $+\delta p$ to form a proof showing $D \mid\sim_L p$. Since $body(r)$ is finite, the proof tree is finite.

If $r \in R_d[p]$, then as before for each a in $body(r)$, $D \mid\sim_L a$. Since $p \in T_D(\mathcal{Y})$, for all $c \in C[p]$ there is a $q \in c - \{p\}$ such that for all rules $s \in R[q]$, $s \prec r$ or else there is a $v \in body(s)$ such that $v \in \mathcal{U}_{\mathcal{Y}}$. By the inductive hypothesis, $D \sim\!|_L v$. Since $\mathcal{U}_{\mathcal{Y}}$ is finite, the number of v's is finite. Adding a tree for each v for each conflict set $c \in C[p]$ as well as adding trees for each $a \in body(r)$ to a root labeled $+\delta p$ forms a proof tree that satisfies definition 3.14.1 (For NDL, it's def. 3.9.1). And so $D \mid\sim_L p$.

(Case 2) Suppose $p \in \mathcal{U}_{\mathcal{X}}$. Then there exists a $\mathcal{Y}$ that is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$. Furthermore, $\mathcal{U}_{\mathcal{X}}$ is an unfounded set *wrt* D and $\mathcal{Y}$.

Let τ_0 be the tree consisting of a single unmarked node labeled $-\delta p$. From τ_0, we construct a series of trees. Given a tree τ_i, we form a new tree τ_{i+1} by picking any unmarked node $x \in \tau_i$ labeled $-\delta q$ for some q such that $q \in \mathcal{U}_{\mathcal{X}}$. Since $q \in \mathcal{U}_{\mathcal{X}}$ and $\mathcal{U}_{\mathcal{X}}$ is an unfounded set *wrt* D and $\mathcal{Y}$, for each rule $r \in R_{sd}[q]$, there is literal $a \in body(r)$ such that either (a) $a \in \mathcal{U}_{\mathcal{Y}}$ or (b) $a \in \mathcal{U}_{\mathcal{X}}$ or (c) $r \in R_d$ and there exists a conflict set $c \in C[q]$ such that for all $u \in c - \{q\}$, there is a rule s such that $body(s) \subseteq \mathcal{T}_{\mathcal{Y}}$ and s is strict or $r \prec s$ (for NDL, $s \not\prec r$). We consider each case in turn.

(2.a) $a \in \mathcal{U}_{\mathcal{Y}}$. By inductive hypothesis $D \sim\!|_L a$. We may append a proof tree τ_a with root $-\delta a$ to node x and mark each node of τ_a.

(2.b) $a \in \mathcal{U}_{\mathcal{X}}$. If x does not already have a child labeled $-\delta a$, then append to x a node y labeled $-\delta a$. If y satisfies condition 2 or 3 in Definition 3.14 (For NDL, 3.9), then mark y. Otherwise, leave y unmarked.

(2.c) Since for all $u \in c - \{q\}$, there is a rule s such that $body(s) \subseteq \mathcal{T}_{\mathcal{Y}}$, it follows by the inductive hypothesis for each $v \in body(s)$ a proof tree τ_v exists with root $+\delta v$. We may append these to node x and mark all nodes of τ_v.

After applying one of the cases 2.a–2.c for each rule $r \in R_{sd}[q]$, examine the resulting tree to see if there is an unmarked non-leaf node z in the tree such that all the children of z are marked. If such a node z is found, mark it. Repeat this procedure until there are no more unmarked nodes in the tree all of whose children are marked. The resulting tree is τ_{i+1}.

$$\text{Let } \tau = \bigcup_{i=0}^{\infty} \tau_i.$$

Suppose x is a marked node in τ. If x was added to τ using case 2.a, then x occurs within a subtree of τ that is a proof tree. So x must satisfy one of the conditions in Definition 3.14(3.9). Similarly, if x was added using case 2.c, then x is part of a valid proof tree and so satisfies the proof conditions. If x was added to τ and marked according to case 2.b, then x is a leaf node in τ and x satisfies condition 2 or 3 of Definition 3.14(3.9). Otherwise, x is a non-leaf node in τ, x was added to τ using condition 2.b, and x was marked because all of its children were marked. Looking at the cases used to add the children of x to τ (we add a child for each rule for q), we see that x must satisfy condition 2 in Definition 3.14(3.9). So if τ is finite and if every node in τ is marked, then τ is a proof tree.

Since cases 2.a–2.c append to node x nodes labeled with a literal from $\mathcal{X}$ or $\mathcal{Y}$ and both of these are finite, it must be the case that the branching factor of τ is finite. So if τ is infinite, then τ must have an infinitely long branch. Consider such a branch. Every node in this branch (other than the top node) must have been added using case 2.b since all the other branches add proof trees which are finite. So every node in the branch must be labeled $-\delta q$ for some literal q. Furthermore, no node in the branch satisfies condition 3 in Definition 3.14 since if it did, it would have been marked when it was added to τ and it would therefore have no children. But since $\mathcal{X}$ is finite, only finitely many literals occur in $\mathcal{X}$. So there must be some literal q such that two different nodes in

our infinite branch are labeled $-\delta q$. But then one of these two nodes does satisfy condition 3 of Definition 3.14(3.9), which is a contradiction. Therefore, τ is not infinite.

Since τ is not infinite, we can let j be a non-negative integer such that $\tau = \tau_j$. Suppose τ_j has an unmarked node. Since a node must be marked if all its children are marked, τ_j must have an unmarked leaf node x. This node must have been added by case 2.b of our construction, and so we can let q be a literal such that x is labeled $-\delta q$, and $q \in \mathcal{U}_{\mathcal{X}}$. Since x is not marked, it satisfies neither condition 2 or 3 of Definition 3.14(3.9). If there is no rule $r \in R_{sd}[q]$, then x satisfies condition 2 of Definition 3.14(3.9). So there is a rule $r \in R_{sd}[q]$, and one of the cases 2.a–2.c applies to x. So there must be some $m > j$ such x has a child node in τ_m. Then x is not a leaf node in τ_m and x is not a leaf node in τ, a contradiction. Therefore, every node in τ satisfies some condition in Definition 3.14(3.9) and τ is a proof tree.

If α is a limit ordinal, then by Lemma 4.21, there is a $\beta < \alpha$ such that $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\beta}$. By the inductive hypothesis, if $p \in \mathcal{T}_{\mathcal{X}}$, then $D \mid\sim_L p$, and if $p \in \mathcal{U}_{\mathcal{X}}$, then $D \sim\!\mid_L p$. □

4.2.3 SOUNDNESS

The proof systems for NDL and ADL are sound *wrt* their counterpart semantic; we do not need to appeal to finite grounded components to show this. However, the below soundness theorem shows that if a proof or refutation tree exists for a literal p, then p appears well–founded or unfounded in some finite grounded component. Given this and the previous result, we conclude that the notion of finite grounded component exactly captures provability and refutability in the proof systems.

The below definition makes the proof of soundness slightly shorter.

Definition 4.26 (Defeat at a Node)**.** *Let $D = \langle R, C, \prec \rangle$ be a defeasible theory, τ a proof tree over D, and n a node of τ. Let $r \in R_d[q]$. A conflict set $c \in C[q]$* defeats *r at n in ADL (NDL) if and only if for all $u \in c - \{q\}$, there is a rule $s \in R[u]$ such that $body(s)$ succeeds at n and s is strict or $r \prec s$ (for NDL, the condition is simply $s \not\prec r$).*

Theorem 4.27 (Soundness)**.** *Let D be a defeasible theory and L one of ADL or NDL. If $D \mid\sim_L p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_{\mathcal{X}}$. If $D \sim\!\!\mid_L p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{U}_{\mathcal{X}}$.*

Proof. If $D \mid\sim_L p$ or $D \sim\!\!\mid_L p$, then there is a proof tree τ showing this. We induct on the depth of τ.

(Base case) Suppose τ is just a single node n labeled $+\delta p$ or $-\delta p$. We consider each case separately.

(Case 1) Suppose that n is labeled $+\delta p$. Then either 3.14.1.a or 3.14.1.b of the definition of proof obtains. If it's 3.14.1.a, since n has no children, there must be a rule $r \in R_s[p]$ such that $body(r) = \varnothing$. By definition 4.8 (Immediate Consequence), $p \in T_D(\mathcal{I}_0)$. But $T_D(\mathcal{I}_0) = \mathcal{T}_1$, and so $p \in \mathcal{T}_1$. Since $\{\}$ is unfounded *wrt* D and $\mathcal{I}_0$ and $p \in T_D(\mathcal{I}_0)$, the interpretation $\langle\{p\}, \{\}\rangle$ forms a finite grounded component of $\mathcal{I}_1$.

If 3.14.1.b obtains, then there is some rule $r \in R_d[p]$ that succeeds at n. Since n has no children, $body(r) = \varnothing$. Let $c \in C[p]$. Since definition 3.14.1.b is satisfied, then there is some $q \in c - \{p\}$ such that for every rule $s \in R[q]$, $body(s)$ fails at n or else $s \prec r$. Since τ consists of a single node, $body(s)$ cannot fail and so $s \prec r$. Generalizing on c, each conflict set in $c \in C[p]$ contains a $q \neq p$ such that for every rule $s \in R[q]$, $s \prec r$. By definition 4.8, $p \in T_D(\mathcal{I}_0) = \mathcal{T}_1$. As before, the interpretation $\langle\{p\}, \{\}\rangle$ forms a finite grounded component of $\mathcal{I}_1$.

(Case 2) Suppose that n is labeled $-\delta p$. Since τ consists of only a single node, failure-by-looping cannot apply and there can be no strict rules with head p. Therefore 3.14.2.b must obtain. Let $r \in R_d[p]$. Since node n has no children, $body(r) = \varnothing$ and so 3.14.2.b.ii must hold. Let $c \in C[p]$ and $q \in c - \{p\}$, and let $s \in R[q]$ such that $body(s)$ succeeds and s is strict or else $r \prec s$ (for NDL, this restriction is relaxed to $s \not\prec r$). Such an s must exist since 3.14.2.b.ii holds. Since τ has only a single node, $body(s) = \varnothing$ and so $body(s) \subseteq \mathcal{T}_0$. Generalizing on q, for every $q \in c - \{p\}$, there exists a rule $s \in R[q]$ such that $body(s) \in \mathcal{T}_0$ and s is strict or else $r \prec s$ (again, for NDL, $s \not\prec r$). Generalizing on r, every rule for p satisfies definition 4.1.2.b (for NDL, 4.2.2.b), and so the unit set

$\{p\}$ is an unfounded set *wrt* to D and $\mathcal{I}_0$. The interpretation $\langle\{\}, \{p\}\rangle$ thus forms a finite grounded component of $\mathcal{I}_1$.

(Induction) Suppose for all $j \leq k$, if τ has depth j and its root is labeled $+\delta p$ ($-\delta p$), then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_{\mathcal{X}}$ ($p \in \mathcal{U}_{\mathcal{X}}$). Suppose τ has depth $k+1$.

(Case 1) Suppose the root n of τ is labeled $+\delta p$. Then 3.14.1.a or 3.14.1.b again holds. If 3.14.1.a holds, then there is a strict rule $r \in R_s[p]$ such that $body(r)$ succeeds at n. For all $q \in body(r)$, there is a child of m labeled $+\delta q$. Each such q is the root of a valid argument tree of maximum depth k, and so by the inductive hypothesis, there exists a finite α_q and a finite grounded component $\mathcal{X}_q$ of $\mathcal{I}_{D,\alpha_q}$ such that $q \in \mathcal{T}_{\mathcal{X}_q}$. From Lemma 4.20, there exists a least $\beta \geq 0$ such that for each $q \in body(r)$, $\mathcal{X}_q$ is a finite grounded component of $\mathcal{I}_{D,\beta}$. Since each α_q is finite, β must be finite as well. Let $\mathcal{X}$ be the least upper bound of these $\mathcal{X}_q$'s. Since $body(r)$ is finite, $\mathcal{X}$ is finite. From Lemma 4.22, $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\beta}$. Thus, we have $body(r) \subseteq \mathcal{T}_{\mathcal{X}}$ and $r \in R_s$. By Definition 4.8.1, $p \in T_D(\mathcal{X})$. Since $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\beta}$ and $p \in T_D(\mathcal{X})$, by monotonicity of T_D, $p \in T_D(\mathcal{I}_{D,\beta})$. The interpretation $\langle\{p\}, \{\}\rangle$ thus forms a finite grounded component of $\mathcal{I}_{\beta+1}$.

If 4.1.1.b holds, then there is a defeasible rule $r \in R_d[p]$ such that $body(r)$ succeeds at n. As before, for each $q \in body(r)$, we have a finite α_q and finite grounded component $\mathcal{X}_q$ of α_q such that $q \in \mathcal{T}_{\mathcal{X}_q}$. Let $c \in C[p]$. Since 4.1.1.b holds, there is a $u \in c - \{p\}$ such that for all rules $s \in R[u]$, either $body(s)$ fails at n or else $s \prec r$. Suppose $body(s)$ fails at n. By definition of failure, there exists a child m of n labeled $-\delta v$, where $v \in body(s)$. Node m is thus the head of a valid proof tree of depth $\leq k$, and so by the inductive hypothesis, there exists a finite α_v and finite grounded component $\mathcal{X}_v$ of α_v such that $v \in \mathcal{U}_{\mathcal{X}_v}$. Generalizing on s and then c, for every $c \in C[p]$, there is a $u \in c - \{p\}$ such that if $s \in R[u]$, then $body(s) \cap \mathcal{U}_{\mathcal{X}_v} \neq \varnothing$ for some finite grounded component $\mathcal{X}_v$ of $\mathcal{I}_{D,\alpha_v}$, or else $s \prec r$.

From Lemmas 4.20, there exists a least $\beta \geq 0$ such that each $\mathcal{X}_q \sqsubseteq \mathcal{I}_{D,\beta}$ and $\mathcal{X}_v \sqsubseteq \mathcal{I}_{D,\beta}$. Let $\mathcal{X}$ be the least upper bound of these $\mathcal{X}_q$'s and $\mathcal{X}_v$'s. Since τ is finite, $\mathcal{X}$ is finite. Also, since each α if finite, β is finite. From Lemma 4.22, $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\beta}$.

Thus, $body(r) \subseteq \mathcal{T}_{\mathcal{X}}$ and for every $c \in C[p]$, there is a $q \in c - \{p\}$ such that if $s \in R[q]$, then $body(s) \cap \mathcal{U}_{\mathcal{X}} \neq \varnothing$ or else $s \prec r$. By definition of 4.8.2, $p \in T_D(\mathcal{X})$. The interpretation $\langle \{p\}, \{\} \rangle$ again forms a finite grounded component of $\mathcal{I}_{\beta+1}$.

(Case 2) Suppose the root n of τ is labeled $-\delta p$. Any branch of a proof tree involving failure-by-looping need not extend beyond the topmost node where definition 3.14.3 (failure-by-looping) applies. As this is so, the tree can be trimmed to that point, and so 3.14.3 only applies to the leaves of the tree. We may assume without loss of generality that τ is of this form.

Define N to be the set of nodes of τ labeled with $-\delta u$ for any u, and S to be the set of the u's. Let n be any node in N. Then n is labeled $-\delta q$ for some q. Node n is either a leaf or an internal node. We treat each case separately.

(Case 2.a) If n is an internal node, then 3.14.2 obtains. Suppose $r \in R_{sd}[q]$. If $r \in R_s[q]$, then $body(r)$ fails at n. By definition of failure, n has a child m labeled $-\delta v$, where $v \in body(r)$. By definition of N and S, $m \in N$ and $v \in S$.

Let $r \in R_d[q]$. Since 3.14.2 holds at n, either (i) the $body(r)$ fails at n and so there is a $a \in body(r)$ and a child m of n such that m is labeled $-\delta a$ and $m \in N$ (and so $a \in S$), or (ii) there is a conflict set $c \in C[q]$ that defeats r at n. If (ii) holds, then for any $u \in c - \{q\}$, there is a $s \in R[u]$ such that $r \prec s$ or s is strict (for NDL, $s \not\prec r$) and for each $v \in body(s)$, there is a valid subtree of n with root labeled $+\delta v$ that constitutes a valid proof. This subtree has depth $\leq k$ and so by the inductive hypothesis there exists a finite $\alpha_v \geq 0$ and finite grounded component $\mathcal{X}_v$ of $\mathcal{I}_{D,\alpha_v}$ such that $v \in \mathcal{T}_{\mathcal{X}_v}$. Generalizing on $v \in S$ and then u, since τ is finite, then by Lemmas 4.20 and 4.22, there exists a least ordinal β and single finite grounded component $\mathcal{X}_r$ of $\mathcal{I}_{D,\beta}$ such that for each $u \in c - \{q\}$, there exists a $s \in R[u]$ such that $body(s) \subseteq \mathcal{T}_{\mathcal{X}_r}$ and s is strict or $r \prec s$ (for NDL, $s \not\prec r$).

Observe that there is a $\mathcal{X}_r$ for each rule $r \in R_{sd}[q]$. $R_{sd}[q]$ might be infinite in size, but since τ is finite, we are assured that the number of $\mathcal{X}_r$'s is finite. Each $\mathcal{X}_r$ is a finite grounded component of some $\mathcal{I}_{D,\beta}$. Let λ be the maximum of the β's and $\mathcal{X}$ the least upper bound of the $\mathcal{X}_r$'s. From Lemmas 4.20 and 4.22 $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\lambda}$. Thus, generalizing on r, (1) for

each $r \in R_s[q]$, $body(r) \cap S \neq \varnothing$, and (2) for each $r \in R_d[q]$ either $body(r) \cap S \neq \varnothing$, or else there is a conflict set $c \in C[q]$ such that for all $u \in c - \{q\}$ there is a rule $s \in R[u]$ such that $body(s) \subseteq \mathcal{T}_{\mathcal{X}}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$).

(Case 2.b) Suppose that n is a leaf node. Then either 3.14.2 or 3.14.3 obtains. If 3.14.2 obtains, then as was shown in the base case, there are no strict rules with head q, and every defeasible rule r is defeated at n by some conflict set c. $\langle\{\}, \{q\}\rangle$ is a finite grounded component of $\mathcal{I}_{D,1}$. If 3.14.3 obtains, then there is a non-leaf node labeled $-\delta q$, and we have shown there that (1) for each $r \in R_s[q]$, $body(r) \cap S \neq \varnothing$, and (2) for each $r \in R_d[q]$ either $body(r) \cap S \neq \varnothing$, or else there is a conflict set $c \in C[q]$ such that for all $u \in c - \{q\}$ there is a rule $s \in R[u]$ such that $body(s) \subseteq \mathcal{T}_{\mathcal{X}}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$).

Given the above 2 cases, by definition S is unfounded with respect to D and $\mathcal{X}$. Note that S is a finite set. As this is so, $\langle\{\}, S\rangle$ forms a finite grounded component with respect to $\lambda + 1$. □

Recall that finite grounded components are closed under union and furthermore, if $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\beta}$ for all $\beta > \alpha$. From this it follows that the finite grounded components constructible from a given proof tree τ can be combined into a single finite grounded component $\mathcal{Z} \sqsubseteq \mathcal{I}_{D,\lambda}$ for some ordinal λ. Since there are only a finite number of components constructible from τ and each of these is associated with a finite ordinal, it must be the case that λ is finite. This in not surprising, but it does show that the proof systems of ADL and NDL must be incomplete *wrt* their counterpart semantics, for we have already seen cases where the well-founded model is reached only at ω or beyond.

Also, since the existence of a proof tree for $+\delta p$ (alternatively, $-\delta p$) implies the existence of a finite grounded component $\mathcal{Z}$ of $\mathcal{I}_{D,\lambda}$ where λ is finite, this implies that all finite grounded components of any interpretation occur first at some finite ordinal. We state this as a corollary.

Corollary 4.28. *Let D be a defeasible theory and L one of NDL or ADL. If $D \mid\!\sim_L p$ ($D \sim\!|_L p$), then there exists a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,n}$ for some finite $n > 0$ such that $p \in \mathcal{T}_{\mathcal{X}}$ ($p \in \mathcal{U}_{\mathcal{X}}$).*

4.3 Locally Finite Theories

The notion of finite grounded component exactly captures when proof and refutation trees exist for literals. However, it is unknown at this point how difficult it is to determine whether a finite grounded component exists for a given literal. For finite theories, the matter is simple, since the well-founded model is itself a finite grounded component. But for arbitrary theories, determining whether an interpretation forms a finite grounded component does not appear to be even recursively enumerable. We have not looked into the matter sufficiently.

Nevertheless, the concept does appear to be of some use. In this section, motivated by a discussion with David Billington, we define the literals that are *locally finite* in a given theory D. We show that if one of these literals is well-founded or unfounded in D, then it is also a member of a finite grounded component (that is, it is finitely well–founded or unfounded). Given the completeness result of the previous section, we may infer that corresponding proof trees exist for such literals.

Definition 4.29. *Let D be a defeasible theory. If $p \in Lit_D$, then $Pre(p)$ is the smallest set such that*

(1) $p \in Pre(p)$, and
(2) for each $q \in Pre(p)$, if $c \in C[q]$, then $c \subseteq Pre(p)$, and
(3) if $q \in Pre(p)$ and $r \in R[q]$, then for each $a \in body(r)$ and each $c \in C[a]$, $c \in Pre(p)$.

Observe that in the third condition, if $q \in Pre(p)$ and $r \in R[q]$, then $body(r) \subseteq Pre(p)$.

Lemma 4.30. *If $Pre(p)$ is locally finite, then for all $q \in Pre(p)$, $Pre(q)$ is locally finite.*

Proof. It is clear from the definition of $Pre(p)$ that $Pre(q) \subseteq Pre(p)$, and so if $Pre(p)$ is finite, $Pre(q)$ must be finite as well. □

Definition 4.31 (Locally Finite Literal)**.** *Let D be a defeasible theory and $p \in Lit_D$. p is* locally finite *in D* iff *$Pre(p)$ is finite.*

Definition 4.32 (Locally Finite Theory)**.** *Let D be a defeasible theory. D is locally finite* iff *for each $p \in Lit_D$, p is locally finite in D.*

To determine whether a literal p is locally finite we start with a queue containing p alone. We then repeat the following process: (1) remove the first element q from the front of the queue; (2) find a rule $r \in R_D[q]$ add $body(r)$ to the queue and delete r from R_D; (3) find a set $c \in C_D[q]$ add it to the queue and delete it from C_D; (4) add q to the end of the queue. Only one occurrence of a literal should be allowed in the queue at any point. If we reach a point at which neither 2 nor 3 are applicable, we put q back into the queue and stop. In that case, p is locally finite. Given that the sets of rules and conflict sets of D are countable, we infer that locally finite literals are at least recognizable.

Returning to Example 4.15, it can be seen that $Pre(p) = \{p, \neg p\} \cup \{q_n | n \in \mathbb{N}\}$ and so p (and consequently D) is not locally finite. Recall that though p is well-founded, no proof tree for $+\delta p$ exists in either NDL or ADL. However, consider the modified theory below.

$\{\} \rightarrow p$
$\{q_0\} \Rightarrow \neg p$
$\{q_{n+1}\} \Rightarrow q_n$ (for each $n \in \mathbb{N}$)

We have made the first rule strict rather than defeasible. A proof tree for $+\delta p$ now exists, and the interpretation $\langle \{p\}, \{\} \rangle$ is a finite grounded component of $\mathcal{I}_{D,1}$. However, p is still not locally finite according to the above definition. This shows that the notion of finite grounded components is in a sense more general than that of locally finite.

The proof that well–founded or unfounded locally finite literals also appear in a finite grounded component is shown below.

Lemma 4.33. *Let L be one of NDL or ADL, D a propositional defeasible theory, and p a locally finite literal in D. For all $\alpha \geq 0$, if $p \in \mathcal{T}_{D,\alpha}$ then there exists a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_X$, and if $p \in \mathcal{U}_{D,\alpha}$ then there exists a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{U}_X$.*

Proof. We induct on α. Suppose the claim holds for all $\beta < \alpha$.

(Case 1) Suppose $p \in \mathcal{T}_{D,\alpha}$ and α is a successor ordinal. Then there exists a rule $r \in R_{sd}[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\alpha-1}$. By inductive hypothesis, for each $q \in body(r)$, there exists a finite grounded component $\mathcal{Y}_q$ of $\mathcal{I}_{D,\alpha-1}$ such that $q \in \mathcal{T}_{\mathcal{Y}_q}$.

Suppose r is strict. Since $body(r)$ is finite and by Lemma 4.22 finite grounded components are closed under finite union, there exists a single finite grounded component $\mathcal{Y}$ of $\mathcal{I}_{D,\alpha-1}$ such that $body(r) \subseteq \mathcal{T}_{\mathcal{Y}}$. Since r is strict, $p \in T_D(\mathcal{Y})$. Since $\mathcal{Y} \sqsubseteq \mathcal{T}_{D,\alpha-1}$, $p \in \mathcal{T}_{D,\alpha}$. Thus, $\langle\{p\}, \{\}\rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$.

If r is defeasible, then for each $c \in C[p]$, there exists an $a \in c - \{p\}$ such that for all $s \in R[a]$, there exists a $v \in body(s)$ such that $v \in \mathcal{U}_{D,\alpha-1}$ or else $s \prec r$. Observe that each v is a member of $Pre(p)$. If $v \in \mathcal{U}_{D,\alpha-1}$, then by inductive hypothesis, there exists a finite grounded component $\mathcal{Y}_v$ of $\mathcal{I}_{D,\alpha-1}$ such that $v \in \mathcal{U}_{\mathcal{Y}_v}$.

Since p is locally finite, $Pre(p)$ is finite, and since each v is a member of $Pre(p)$, there are only a finite number of v's. Thus we may combine each $\mathcal{Y}_q$ and $\mathcal{Y}_v$ into a single finite grounded component $\mathcal{Y}$ of $\mathcal{I}_{D,\alpha-1}$. As before, $p \in T_D(\mathcal{Y})$. We conclude that $\langle\{p\}, \{\}\rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$.

(Case 2) Suppose $p \in \mathcal{U}_{D,\alpha} \cap Pre(p)$ and let a be any literal in $\mathcal{U}_{D,\alpha} \cap Pre(p)$. Then for each rule $r \in R_s[a]$, there exists a $v \in body(r)$ such that $v \in (\mathcal{U}_{D,\alpha} \cup \mathcal{U}_{D,\alpha-1})$. By monotonicity, $v \in \mathcal{U}_{D,\alpha}$. Since $v \in body(r)$ and $head(r) \in Pre(p)$ it follows that $v \in Pre(p)$.

Suppose $r \in R_d[a]$. Then either (i) as above there exists a $v \in body(r)$ such that $v \in \mathcal{U}_{D,\alpha} \cap Pre(p)$, or else (ii) there exists a $c \in C[a]$ such that for each $q \in c - \{a\}$, there exists an $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_{D,\alpha-1}$ and $s \not\prec r$ (for ADL, s is strict or else $r \prec s$). Note that by definition of $Pre(p)$, $body(s) \subseteq Pre(p)$. If (ii) holds, then by inductive hypothesis, (since $Pre(p)$ is finite) there exists a finite grounded component $\mathcal{Y}_s$ of $\mathcal{I}_{D,\alpha-1}$ such that $body(s) \subseteq \mathcal{T}_{\mathcal{Y}_s}$. Let $\mathcal{Y}_r$ be the least upper bound of the $\mathcal{Y}_s$'s.

Generalizing on r and then on a, let $\mathcal{Y}$ be the least upper bound of the $\mathcal{Y}_r$'s. Since $Pre(p)$ is finite, it follows by Lemma 4.22 that $\mathcal{Y}$ is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$. It can be seen that $Pre(p) \cap \mathcal{U}_{D,\alpha}$ is an unfounded set *wrt* D and $\mathcal{Y}$. As such $\mathcal{X} = \langle\{\}, Pre(p) \cap \mathcal{U}_{D,\alpha}\rangle$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$. Furthermore, $p \in \mathcal{U}_{\mathcal{X}}$.

If α is a limit ordinal, there exists a successor ordinal $\beta < \alpha$ such that $p \in \mathcal{T}_{D,\beta}$ ($p \in \mathcal{U}_{D,\beta}$), and so by the inductive hypothesis, there exists a finite grounded component $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\beta}$ such that

$p \in \mathcal{T}_{\mathcal{X}}$ ($p \in \mathcal{U}_{\mathcal{X}}$). Since $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\beta}$, it is also a finite grounded component of $\mathcal{I}_{D,\alpha}$. □

For a locally finite literal p in D, in order to determine whether p is well–founded or unfounded we need only examine a finite portion of D. Particularly, we need only examine the portion of D defined over the literals appearing in $Pre(p)$. We define this finite sub–theory below.

Definition 4.34. *Let $D = \langle R_D, C_D, \prec_D \rangle$ be a defeasible theory and p a literal locally finite in D. Define D^p to be the smallest theory such that*

(1) $R_{D^p} = \{r | r \in R_D$ *and* $(body(r) \cup head(r)) \subseteq Pre(p)\}$
(2) $C_{D^p} = \{c | c \in C_D$ *and* $c \subseteq Pre(p)\}$
(3) $\prec_{D^p} = \{(r,s) | \{r,s\} \subseteq R_{D^p}\}$

Since p is locally finite in D, $Pre(p)$ is finite (by definition), and so D^p is finite.

Lemma 4.35. *Let D be a defeasible theory and x a locally finite literal of D. For all $p \in Pre(x)$ the following holds:*

(1) $R_{D^x}[p] = R_D[p]$
(2) $C_{D^x}[p] = C_D[p]$

Proof. Suppose $p \in Pre(x)$. Obviously, $R_{D^x}[p] \subseteq R_D[p]$ and $C_{D^x}[p] \subseteq C_D[p]$. In the other direction, if $r \in R_D[p]$, then by definition of $Pre(x)$, $body(r) \in Pre(x)$. Since $body(r) \cup \{p\} \subseteq Pre(x)$, then by definition of D^x, $r \in R_{D^x}$. Similarly, if $c \in C_D[p]$, then $c \subseteq Pre(x)$ and so by definition of D^x, $c \in C_{D^x}$. □

Lemma 4.36. *Let D be a defeasible theory and x a locally finite literal of D. For all $p \in Pre(x)$, $D \mid\!\approx_L p$* iff *$D^x \mid\!\approx_L p$, and $D \approx\!\!\mid_L p$* iff *$D^x \approx\!\!\mid_L p$.*

Proof. **(LR)** Suppose that for all $\beta < \alpha$, if $p \in \mathcal{T}_{D,\beta}$ then $p \in \mathcal{T}_{D^x,\beta}$, and if $p \in \mathcal{U}_{D,\beta}$ then $p \in \mathcal{U}_{D^x,\beta}$. Suppose that α is a successor ordinal. We proceed in cases.

(Case 1) Let $p \in \mathcal{T}_{D,\alpha}$. Then there exists a rule $r \in R_{D,sd}[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\alpha-1}$. Since $p \in Pre(x)$, then by Lemma 4.35, $r \in R_{D^x}$. By inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D^x,\alpha-1}$. If $r \in R_{D^x,s}$, then by definition of T_{D^x}, $p \in T_{D^x}(\mathcal{I}_{D^x,\alpha-1}) = \mathcal{T}_{D^x,\alpha}$.

Suppose $r \in R_{D^x,d}$. Then for all $c \in C_D[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R_D[s]$, either $body(s) \cap \mathcal{U}_{D,\alpha-1} \neq \varnothing$ or else $s \prec r$. If $body(s) \cap \mathcal{U}_{D,\alpha-1} \neq \varnothing$, then by inductive hypothesis, $body(s) \cap \mathcal{U}_{D^x,\alpha-1} \neq \varnothing$. By definition of $Pre(x)$, $C_D[p] = C_{D^x}[p]$ and for each $c \in C[p]$ and $q \in c$, $R_D[q] = R_{D^x}[q]$. By definition of T_{D^x}, $p \in T_{D^x}(\mathcal{I}_{D^x,\alpha-1}) = \mathcal{T}_{D^x,\alpha}$.

(Case 2) Let $p \in \mathcal{U}_{D,\alpha}$ and let a be any literal in $\mathcal{U}_{D,\alpha} \cap Pre(x)$. Then for each $r \in R_{D,s}$, $body(r) \cap (\mathcal{U}_{D,\alpha} \cup \mathcal{U}_{D,\alpha-1}) \neq \varnothing$. By monotonicity of $\mathcal{U}_D$, $body(r) \cap \mathcal{U}_{D,\alpha} \neq \varnothing$.

For each $r \in R_{D,d}$, either $body(r) \cap \mathcal{U}_{D,\alpha} \neq \varnothing$, or else there exists a $c \in C_D[a]$ such that for each $q \in c - \{a\}$ there exists a rule $s \in R_D[q]$ such that $body(s) \subseteq \mathcal{T}_{D,\alpha-1}$ and $s \not\prec r$ (for ADL, s is strict or else $r \prec s$). By inductive hypothesis, $body(s) \subseteq \mathcal{T}_{D^x,\alpha-1}$. For each s, by definition of $Pre(x)$ and D^x, $s \in D^x$. Similarly, since $a \in Pre(x)$, $C_D[a] = C_{D^x}[a]$, and also, for each $r \in R[a]$, $r \in D^x$.

Generalizing on a, for each $a \in \mathcal{U}_{D,\alpha} \cap Pre(x)$, for each $r \in R_{D^x,sd}$ either $body(r) \cap \mathcal{U}_{D,\alpha} \neq \emptyset$, or else $r \in R_{D^x,d}$ and there exists a $c \in C_{D^x}[a]$ such that for each $q \in c - \{a\}$ there exists a rule $s \in R_{D^x}[q]$ such that $body(s) \subseteq \mathcal{T}_{D^x,\alpha-1}$ and $s \not\prec r$ (for ADL, s is strict or else $r \prec s$). $\mathcal{U}_{D,\alpha} \cap Pre(x)$ is thus an unfounded set *wrt* D^x and $\mathcal{I}_{D^x,\alpha-1}$, and so $(\mathcal{U}_{D,\alpha} \cap Pre(x)) \subseteq \mathcal{U}_{D^x,\alpha}$. In particular, $p \in \mathcal{U}_{D^x,\alpha}$.

(RL) Suppose that for all $\beta < \alpha$, if $p \in \mathcal{T}_{D^x,\beta}$ then $p \in \mathcal{T}_{D,\beta}$, and if $p \in \mathcal{U}_{D^x,\beta}$ then $p \in \mathcal{U}_{D,\beta}$. Let α be a successor ordinal. We proceed in cases.

(Case 1) Let $p \in \mathcal{T}_{D^x,\alpha}$. Then either there exists a rule $r \in R_{D^x,s}[p]$ such that $body(r) \subseteq \mathcal{T}_{D^x,\alpha-1}$, or else $r \in R_{D^x,d}$, $body(r) \subseteq \mathcal{T}_{D^x,\alpha-1}$, and for all $c \in C_{D^x}[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R_{D^x}[s]$, either $body(s) \cap \mathcal{U}_{D^x,\alpha-1} \neq \varnothing$ or else $s \prec r$. Clearly we have $R_D[p] = R_{D^x}[p]$, $C_D[p] = C_{D^x}[p]$, and for each $c \in C[p]$ and $q \in c$, $R_D[q] = R_{D^x}[q]$. Thus by inductive hypothesis, we have: either there exists a rule $r \in R_{D,s}[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\alpha-1}$, or else $r \in R_{D,d}$, $body(r) \subseteq \mathcal{T}_{D,\alpha-1}$, and for all $c \in C_D[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R_D[s]$, either $body(s) \cap \mathcal{U}_{D,\alpha-1} \neq \varnothing$ or else $s \prec r$. By definition of $T_D(\mathcal{I}_{D,\alpha-1})$, $p \in \mathcal{T}_{D,\alpha}$.

(Case 2) Let $p \in \mathcal{U}_{D^x,\alpha}$ and let a be any literal in $\mathcal{U}_{D^x,\alpha} \cap Pre(x)$. Then for each $r \in R_{D^x,s}$, $body(r) \cap (\mathcal{U}_{D^x,\alpha} \cup \mathcal{U}_{D^x,\alpha-1}) \neq \varnothing$. For each $r \in R_{D,d}$, either $body(r) \cap (\mathcal{U}_{D^x,\alpha} \cup \mathcal{U}_{D^x,\alpha-1}) \neq \varnothing$,

or else there exists a $c \in C_{D^x}[a]$ such that for each $q \in c - \{a\}$ there exists a rule $s \in R_{D^x}[q]$ such that $body(s) \subseteq \mathcal{T}_{D^x,\alpha-1}$ and $s \not\prec r$ (for ADL, s is strict or else $r \prec s$).

By definition of $Pre(x)$ and D^x, it is clear that for each $a \in \mathcal{U}_{D^x,\alpha} \cap Pre(x)$, we have $R_D[a] = R_{D^x}[a]$ and $C_D[a] = C_{D^x}[a]$, and for each $c \in C_{D^x}[a]$ and each $q \in c - \{a\}$, we have $R_D[q] = R_{D^x}[q]$. Thus by inductive hypothesis we have: For each $a \in \mathcal{U}_{D^x,\alpha} \cap Pre(x)$, for each $r \in R_{D,s}$, $body(r) \cap ((\mathcal{U}_{D^x,\alpha} \cap Pre(x)) \cup \mathcal{U}_{D,\alpha-1}) \neq \varnothing$. For each $r \in R_{D,d}$, either $body(r) \cap ((\mathcal{U}_{D^x,\alpha} \cap Pre(x)) \cup \mathcal{U}_{D,\alpha-1}) \neq \varnothing$, or else there exists a $c \in C_D[a]$ such that for each $q \in c - \{a\}$ there exists a rule $s \in R_D[q]$ such that $body(s) \subseteq \mathcal{T}_{D,\alpha-1}$ and $s \not\prec r$ (for ADL, s is strict or else $r \prec s$).

Generalizing on a, it can be seen that $\mathcal{U}_{D^x,\alpha} \cap Pre(x)$ is an unfounded set *wrt* D and $\mathcal{I}_{D,\alpha-1}$, and so $(\mathcal{U}_{D^x,\alpha} \cap Pre(x)) \subseteq \mathcal{U}_{D,\alpha}$. In particular, $p \in \mathcal{U}_{D,\alpha}$. □

4.4 Failure of Cut and Cautious Monotony

NDL and ADL are not as well-behaved as we would like. Specifically, if no special restrictions are made on defeasible theories, then the logics do not satisfy Cut and Cautious Monotony. However, they do satisfy these properties provided that $\prec$ is transitive. Proofs of this and lesser claims are given in the next section. In this section, we provided examples showing Cut and Cautious Monotony fail in general.

Recall the versions of Cut and Cautious Monotony provided in chapter 1.

(1) **Cut**: If $\Gamma \mathrel{|\!\sim} p$ and $\Gamma \cup \{p\} \mathrel{|\!\sim} q$, then $\Gamma \mathrel{|\!\sim} q$.
(2) **Cautious Monotony**: If $\Gamma \mathrel{|\!\sim} p$ and $\Gamma \mathrel{|\!\sim} q$, then $\Gamma \cup \{p\} \mathrel{|\!\sim} q$.

There, no restriction was placed upon what p and q might be (other than that they are propositions). However, observe that a defeasible theory consists of a set of rules and that the consequences of the theory are a set of literals. And so, if we are to add a consequence to the theory, we must add it as either a strict or a defeasible rule (more specifically, as a fact or a presumption). In the following, we will consider what happens in either case.

Example 4.37.

(1) $\{\} \Rightarrow q$
(2) $\{s\} \Rightarrow p$
(3) $\{\} \Rightarrow r$
(4) $\{q, r\} \rightarrow \neg p$
(5) $\{\} \rightarrow s$
(6) $\{\} \rightarrow t$
(7) $\{r\} \rightarrow \neg t$

$2 \prec 1$, $3 \prec 2$

The relevant extended conflict sets of the above theory are $\{p, \neg p\}$, $\{t, \neg t\}$, $\{p, q, r, t\}$, and $\{r, t\}$. Under both NDL and ADL, since rule 1 is supported and rule 2 is inferior to 1, we can derive q. r is refuted because the strict rule 6 defeats rule 3 (the only rule for r). Since r is refuted, rule 4 fails; furthermore, rule 3 is inferior to rule 2, and so p is defeasibly provable in both NDL and ADL. The well-founded model according to both is $\langle\{p, q, s, t\}, \{\neg p, r, \neg t\}\rangle$.

Since p is derivable from the theory anyway, suppose that we add it to the theory as a fact (*i.e.* the strict rule 8 with empty body) or as a presumption (i.e., $\{\} \Rightarrow p$). The new theory is:

Example 4.38.

(1) $\{\} \Rightarrow q$
(2) $\{s\} \Rightarrow p$
(3) $\{\} \Rightarrow r$
(4) $\{q, r\} \rightarrow \neg p$
(5) $\{\} \rightarrow s$
(6) $\{\} \rightarrow t$
(7) $\{r\} \rightarrow \neg t$
(8) $\{\} \rightarrow p$ *(or* $\{\} \Rightarrow p$*)*

$2 \prec 1$, $3 \prec 2$

Reconsidering literal q, we see that though rule 1 is still supported, there is now no literal $v \in \{p, q, r\} - \{q\}$ such that all rules for v fail or are inferior to rule 1. And so q cannot be derived in either NDL or ADL. According to NDL, the well-founded model of the modified theory is $\langle\{p, s, t\}, \{\neg p, q, r, \neg t\}\rangle$. According to ADL, its well-founded model is $\langle\{p, s, t\}, \{\neg p, r, \neg t\}\rangle$. Importantly, the same results are obtained whether p is added as a fact or as a presumption. These

results are significant, since they mean that both NDL and ADL (at least when priorities are used) fail to satisfy Cautious Monotony. We cannot add the consequences of a theory as premises without potentially eliminating other consequences.

4.4.1 The Failure of Cut

Example 4.39.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $\{\} \Rightarrow r$
(4) $\{q, r\} \rightarrow \neg p$
(5) $\{r\} \Rightarrow t$
(6) $\{\} \Rightarrow \neg t$

$1 \prec 3$, $2 \prec 1$, $6 \prec 5$

The conflict sets are $\{p, \neg p\}$, $\{q, \neg q\}$, $\{r, \neg r\}$, $\{t, \neg t\}$, and $\{p, q, r\}$. Here, q is refuted under NDL because its only rule (2) is not superior to either 1 or 3. Given this, rule 4 fails. Since rule 4 fails and rule 2 is inferior to rule 1, we can derive p. Since rule 1 is inferior to 3, we can derive r. Since 5 is then supported and superior to 6, we conclude t. $\neg t$ is refuted. However, if we add p:

Example 4.40.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $\{\} \Rightarrow r$
(4) $\{q, r\} \rightarrow \neg p$
(5) $\{r\} \Rightarrow t$
(6) $\{\} \Rightarrow \neg t$
(7) $\{\} \rightarrow p$

$2 \prec 1$, $1 \prec 3$, $6 \prec 5$

Since $\{p, q, r\}$ is a conflict set and both rule 2 and 7 are supported and not inferior to 3, r is refuted. Since r is refuted, rule 5 fails. This allows rule 6 to fire, and so $\neg t$ is derivable. This entails that NDL fails Cut: $D \mathrel{|\!\sim} p$ and $D \cup \{\{\} \rightarrow p\} \mathrel{|\!\sim} \neg t$, but $D \not\mathrel{|\!\sim} \neg t$. The example also works if rule 7 is defeasible rather than strict.

The below examples shows that Cut fails for ADL as well, provided that it is $\{\} \rightarrow p$ that is added to D.

Example 4.41.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $\{\} \Rightarrow r$
(4) $\{q, r\} \rightarrow \neg p$
(5) $\{q\} \Rightarrow t$
(6) $\{\} \Rightarrow \neg t$
(7) $\{\} \Rightarrow \neg r$

$3 \prec 1$, $2 \prec 3$, $3 \prec 1$, $6 \prec 5$, $3 \prec 7$.

The conflict sets are $\{p, \neg p\}$, $\{q, \neg q\}$,$\{t, \neg t\}$, $\{r \neg r\}$, and $\{p, q, r\}$. Here, r is refuted because there is a rule for $\neg r$ (namely rule 7) that is supported and superior to rule 3. Because of this, rule 4 fails, and since rule 1 is superior to rule 3, we derive p. q, however, cannot be derived because there is a conflict set $\{p, q, r\}$ and rules (1,3), are both supported and not inferior to rule 2 (the only rule for q). q is not refuted, either, since rule 1 is not superior to 2. Because of this, t is ambiguous, and neither t nor $\neg t$ can be derived or refuted.

However, if we add p as a fact (*i.e.*, as $(8)\{\} \rightarrow p$), then p is obviously derivable via the strict rule 8. Rule 3 is superior to rule 2. Because of these two rules, q is refuted. Since q is refuted, rule 5 fails. This allows rule 6 to fire, and so $\neg t$ is derivable. Since $D \not\sim_{ADL} p$ and $D \cup \{\{\} \rightarrow p\} \sim_{ADL} \neg t$ but $D \not\sim_{ADL} \neg t$, ADL violates Cut.

The example will not serve as a counterexample if rule 8 is defeasible, however. Indeed, ADL does satisfy Cut if it is $\{\} \Rightarrow p$ that is added to a theory and not $\{\} \rightarrow p$. This is shown in Appendix C.

4.5 Success for Theories with Transitive Priorities

Cautious Monotony and Cut fail in general in the context of prioritized theories for the reason that the information contained in the priority relation is not preserved when we add $\{\} \Rightarrow p$ or $\{\} \rightarrow p$

as a premise to the theory. When the precedence relation is transitive, however, then ADL and NDL satisfy both properties. We now prove that this is so. We actually prove somewhat more general claims: If $D \not\approx_L p$ and $\prec$ transitive, then we may add either $\{\} \Rightarrow p$ or $\{\} \rightarrow p$ to D and not affect the consequences. That is, it doesn't matter if we add p as a fact or a presumption. As a special case, the properties hold for unprioritized theories.

4.5.1 CAUTIOUS MONOTONY

Lemma 4.42. *Let L be one of NDL or ADL and $D = \langle R_D, C_D, \prec \rangle$ a defeasible theory such that $\prec$ is transitive. Let $E = \langle R_D \cup X, C_D, \prec \rangle$ where $X = \{\{\} \rightarrow p\}$ or $X = \{\{\} \Rightarrow p\}$.* For all $\kappa \geq 0$, if $p \in \mathcal{T}_{D,\kappa}$ and $q \in \mathcal{T}_{D,\kappa}$, then $q \in \mathcal{T}_{E,WF}$. For all $\kappa \geq 0$, if $p \in \mathcal{T}_{D,\kappa}$ and $q \in \mathcal{U}_{D,\kappa}$, then $q \in \mathcal{U}_{E,WF}$.

Proof. Suppose the above holds for all $\kappa < \lambda$. The hypothesis is trivially satisfied for $\kappa = 0$. Note that we may not assume $q \neq p$. We proceed in cases:

(Case 1) Suppose that $q \in \mathcal{T}_{D,\lambda}$ and $p \in \mathcal{T}_{D,\lambda}$ and that λ is a successor ordinal. Then there exists a strict or defeasible rule $r_q \in R_D[q]$ such that $body(r_q) \subseteq \mathcal{T}_{D,\lambda-1}$. Since $R_D[q] \subset R_E[q]$, $r_q \in R_E[q]$. By the inductive hypothesis, $body(r_q) \subseteq \mathcal{T}_{E,WF}$. If r_q is strict, then clearly $q \in \mathcal{T}_{E,WF}$ by definition of T_E and $\mathcal{I}_{E,WF}$.

Suppose r_q is defeasible and let $c_1 \in C[q]$. Then α below holds:

α: There exists a $w \in c_1 - \{q\}$ such that for all $r_w \in R_D[w]$, either $r_w \prec r_q$ or $body(r_w) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$.

Obviously, either $w = p$ or $w \neq p$. We'll consider each possibility.

(Case 1.1) Suppose $w \neq p$. Then $R_D[w] = R_E[w]$. By the inductive hypothesis and α above, for all $r_w \in R_E[w]$, either $r_w \prec r_q$ or $body(r_w) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

(Case 1.2) If $w = p$, then since $p \in \mathcal{T}_{D,\lambda}$ there exists an $r_p \in R_D[p]$ such that $body(r_p) \subseteq \mathcal{T}_{D,\lambda-1}$ and either r_p is strict, or else r_p is defeasible and there exists a $u \in c_1 - \{p\}$ such that for every rule $r_u \in R[u]$, either $body(r_u) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$ or $r_u \prec r_p$. Since $body(r_p) \subseteq \mathcal{T}_{D,\lambda-1}$, then by coherence $body(r_p) \cap \mathcal{U}_{D,\lambda-1} = \varnothing$. Since $w = p$ and $body(r_p) \cap \mathcal{U}_{D,\lambda-1} = \varnothing$, then from α we have that r_p is

defeasible and $r_p \prec r_q$. Since the precedence relation is acyclic, $r_q \not\prec r_p$. Since $body(r_q) \subseteq \mathcal{T}_{D,\lambda-1}$, then by coherence we have $body(r_q) \cap \mathcal{U}_{D,\lambda-1} = \varnothing$. Since $body(r_q) \cap \mathcal{U}_{D,\lambda-1} = \varnothing$ and $r_q \not\prec r_p$, it must be the case that $q \neq u$. Since the precedence relation is transitive, there is a $u \in c_1 - \{p, q\}$ such that for every rule $r_u \in R_D[u]$, either $body(r_u) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$ or $r_u \prec r_q$. By the inductive hypothesis, for every rule $r_u \in R_D[u]$, either $body(r_u) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or $r_u \prec r_q$. Since $u \neq p$, $R_D[u] = R_E[u]$.

Generalizing on c_1, for each $c \in C_E[q]$, there is a $w \in c - \{q\}$ such that for each $s \in R_E[w]$, $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or else $s \prec r$. Recall that $body(r_q) \subseteq \mathcal{T}_{E,WF}$. By definition of T_E and $\mathcal{I}_{E,WF}$, $q \in \mathcal{T}_{E,WF}$.

If λ is a limit ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{D,\kappa}$ and $p \in \mathcal{T}_{D,\kappa}$. By the inductive hypothesis, $q \in \mathcal{T}_{E,WF}$.

(Case 2) Suppose $p \in \mathcal{T}_{D,\lambda}$ and $q \in \mathcal{U}_{D,\lambda}$ and that λ is a successor ordinal. Let a be any literal such that $a \in \mathcal{U}_{D,\lambda}$. Since $\mathcal{I}_\lambda$ is coherent and $p \in \mathcal{T}_{D,\lambda}$, $a \neq p$ and so $R_D[a] = R_E[a]$. We will consider each $r_a \in R_{D,sd}[a]$.

(Case 2.1) Suppose $r_a \in R_{D,s}[a]$. Then $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$. By the inductive hypothesis, $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{E,WF}) \neq \varnothing$.

(Case 2.2) Suppose $r_a \in R_{D,d}[a]$. Then either $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$, or else there exists a $c \in C_D[a]$ such that for each $w \in c - \{a\}$ there is a rule $r_w \in R_D[w]$ such that $body(r_w) \subseteq \mathcal{T}_{D,\lambda-1}$ and $r_w \not\prec r_a$ (for ADL, $r_a \prec r_w$ or r_w is strict). If the former, then by the inductive hypothesis $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{E,WF}) \neq \varnothing$. Suppose it's the latter. Since $R_D \subset R_E$ and $C_D = C_E$, it follows that $c \in C_E[a]$ and for each $w \in c - \{a\}$ there is an $r_w \in R_E[w]$ such that (by the inductive hypothesis) $body(r_w) \subseteq \mathcal{T}_{E,WF}$ and $r_w \not\prec r_a$ (for ADL, $r_a \prec r_w$ or r_w is strict). In summary, either $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{E,WF}) \neq \varnothing$, or else there exists a $c \in C_E[a]$ such that for each $w \in c - \{a\}$ there is a rule $r_w \in R_E[w]$ such that $body(r_w) \subseteq \mathcal{T}_{E,WF}$ and $r_w \not\prec r_a$ (for ADL, $r_a \prec r_w$ or r_w is strict).

Generalizing on r_a, for each $r_a \in R_{E,s}[a]$, $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{E,WF}) \neq \varnothing$, and for each $r_a \in R_{E,d}[a]$ either $body(r_a) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{E,WF}) \neq \varnothing$, or else there exists a conflict set $c \in C_E[a]$

such that for each $w \in c - \{a\}$ there is a rule $r_w \in R_E[w]$ such that $body(r_w) \subseteq \mathcal{T}_{E,WF}$ and $r_w \not\prec r_a$ (for ADL, $r_a \prec r_w$ or r_w is strict). Generalizing on a, it can be seen that $\mathcal{U}_{D,\lambda}$ satisfies the definition of unfounded set *wrt* E and $\mathcal{I}_{E,WF}$, and so $\mathcal{U}_{D,\lambda} \subseteq U_E(\mathcal{U}_{E,WF})$ (hence $\mathcal{U}_{D,\lambda} \subseteq \mathcal{U}_{E,WF}$). Since $q \in \mathcal{U}_{D,\lambda}$, $q \in \mathcal{U}_{E,WF}$.

If λ is a limit ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{D,\kappa}$ and $p \in \mathcal{T}_{D,\kappa}$. By the inductive hypothesis, $q \in \mathcal{U}_{E,WF}$. □

Theorem 4.43 (Cautious Monotony for Theories with Transitive Priorities)**.** *Let L be one of NDL or ADL and $D = \langle R_D, C_D, \prec \rangle$ a defeasible theory such that $\prec$ is transitive. Let $E = \langle R_D \cup X, C_D, \prec \rangle$ where $X = \{\{\} \rightarrow p\}$ or $X = \{\{\} \Rightarrow p\}$. If $D \mid\!\approx_L p$ and $D \mid\!\approx_L q$, then $E \mid\!\approx_L q$. If $D \mid\!\approx_L p$ and $D \mid\!\not\approx_L q$, then $E \mid\!\not\approx_L q$.*

Proof. Suppose $D \mid\!\approx_L p$ and $D \mid\!\approx_L q$. Then there exists a least $\lambda \geq 0$ such that $p \in \mathcal{T}_{D,\lambda}$ and $q \in \mathcal{T}_{D,\lambda}$. By Lemma 4.42 $q \in \mathcal{T}_{E,WF}$, and so $E \mid\!\approx_L q$.

Now Suppose $D \mid\!\approx_L p$ and $D \mid\!\not\approx_L q$. Then there exists a least $\lambda \geq 0$ such that $p \in \mathcal{T}_{D,\lambda}$ and $q \in \mathcal{U}_{D,\lambda}$. By Lemma 4.42 $q \in \mathcal{U}_{E,WF}$, and so $E \mid\!\not\approx_L q$. □

4.5.2 CUT

Lemma 4.44. *Let L be NDL or ADL, D a defeasible theory such that $D \mid\!\approx_L p$ and $\prec$ is transitive. Let $E = \langle R_D \cup X, C_D, \prec \rangle$ where $X = \{\rightarrow p\}$ or $X = \{\Rightarrow p\}$. For all $\kappa \geq 0$, if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\kappa}$, then $q \in \mathcal{T}_{D,WF}$. If $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{U}_{E,\kappa}$, then $q \in \mathcal{U}_{D,WF}$.*

Proof. Since $\mathcal{I}_{E,0} = \langle \varnothing, \varnothing \rangle$, the above claim is trivially satisfied for $\kappa = 0$. Suppose it holds for all $\kappa < \lambda$. We proceed in cases.

(Case 1) Suppose $D \mid\!\approx_L p$ and $p \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{T}_{E,\lambda}$. Suppose that λ is a successor ordinal. If $q = p$, then by assumption $D \mid\!\approx_L p$ (*i.e.* $p \in \mathcal{T}_{D,WF}$).

Let $q \neq p$. Then there exists a rule $r \in R_E[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. Since $q \neq p$, $r \in R_D[q]$ and by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$. If r is strict, then clearly $q \in \mathcal{T}_{D,WF}$ by definition of T_D and $\mathcal{I}_{D,WF}$.

Suppose r is defeasible. As before, $body(r) \subseteq \mathcal{T}_{D,WF}$. Let $c \in C_E[q]$. Then there exists a $w \in c - \{q\}$ such that for all $s \in R_E[w]$, $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$. Since $t = \{\} \dashrightarrow p \in R_E$ (where $X = \{t\}$) and t is not mentioned in $\prec_E$, it follows that $body(t) \cap \mathcal{U}_{E,\lambda-1} = \varnothing$ and $t \not\prec r$. Thus $w \neq p$, and so $R_E[w] = R_D[w]$. by the inductive hypothesis for all $s \in R_D[w]$, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or else $s \prec r$.

Note that $C_E = C_D$. Generalizing on c, we have $body(r) \subseteq \mathcal{T}_{D,WF}$ and for each $c \in C_D[q]$ there exists a $w \in c - \{q\}$ such that for all $s \in R_D[w]$, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or else $s \prec r$. By definition of T_D and $\mathcal{I}_{D,WF}$, $q \in \mathcal{T}_{D,WF}$.

If λ is a limit ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{E,\kappa}$ and $p \in \mathcal{T}_{E,\kappa}$. By the inductive hypothesis, $q \in \mathcal{T}_{D,WF}$.

(Case 2) Suppose $D \mid\!\approx_L p$ and $p \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{U}_{E,\lambda}$. Suppose that λ is a successor ordinal. Let a be any literal such that $a \in \mathcal{U}_{E,\lambda}$. By coherence, $a \neq p$ and so $R_E[a] = R_D[a]$.

Since $a \in \mathcal{U}_{E,\lambda}$, then for all $r \in R_{E,sd}[a]$, $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$, or else r is defeasible and there exists a $c \in C_E[a]$ such that for each $w \in c - \{a\}$ there is a rule $s_w \in R_E[w]$ such that $body(s_w) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s_w \not\prec r$ (for ADL, s_w is strict or $r \prec s_w$). We'll consider each in turn.

(Case 2.1) Let $r \in R_{E,s}[a]$. Then $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. By the inductive hypothesis, $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$.

(Case 2.2) Let $r \in R_{E,d}$. Then either (1) $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$, or (2) else there exists a $c \in C_E[a]$ such that for each $w \in c - \{a\}$ there is a rule $s_w \in R_E[w]$ such that $body(s_w) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s_w \not\prec r$ (for ADL, s_w is strict or $r \prec s_w$).

(Case 2.2.1) Suppose $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. by the inductive hypothesis, $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$.

(Case 2.2.2) Suppose there exists a $c \in C_E[a]$ such that for each $w \in c - \{a\}$ there is a rule $s_w \in R_E[w]$ such that $body(s_w) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s_w \not\prec r$ (for ADL, s_w is strict or $r \prec s_w$). There are two cases to consider: (1) where $s_w \in R_D$, and (2) where $s_w \notin R_D$.

(1) If $s_w \in R_D$, then by the inductive hypothesis, we have $body(s_w) \subseteq \mathcal{T}_{D,WF}$ and $s_w \not\prec r$ (for ADL, s_w is strict or $r \prec s_w$).

(2) If $s_w \notin R_D$, then $s_w = t' : \{\} \dashrightarrow p$. Since $D \not\approx_L p$, there exists a rule $t \in R_D[p]$ such that $body(t) \subseteq \mathcal{T}_{D,WF}$ and for some $u \in c - \{p\}$, for each $r_u \in R_D[u]$ either $body(r_u) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $r_u \prec t$.

(2.1) Suppose $u = a$. Then $body(t) \subseteq \mathcal{T}_{D,WF}$ and either $body(r) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $r \prec t$; if the latter, then $t \not\prec r$.

(2.2) Suppose $u \neq a$, then $u \in c - \{p, a\}$. Since for each $w \in c - \{a\}$ there is a rule $s_w \in R_E[w]$ such that $body(s_w) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s_w \not\prec r$ (for ADL, s_w is strict or $r \prec s_w$), it follows that there is a particular $s_u \in R_D[u]$ such that $body(s_u) \subseteq \mathcal{T}_{D,WF}$ and $s_u \not\prec r$ (for ADL, s_u is strict or $r \prec s_u$). By coherence $body(s_u) \cap \mathcal{U}_{D,WF} = \varnothing$, and so $s_u \prec t$. Since $s_u \prec t$, s_u must be defeasible (and so for ADL, $r \prec s_u$). However, since the precedence relation is transitive and $s_u \not\prec r$ (for ADL $r \prec s_u$) and $s_u \prec t$, we have $t \not\prec r$ (for ADL, $r \prec t$).

Generalizing on the literals of $c - \{a\}$, for each $w \in c - \{a\}$, there exists a $s_w \in R_D[q]$ such that $body(s_w) \subseteq \mathcal{T}_{D,WF}$ and either $body(r) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $s_w \not\prec r$ (for ADL, s_w is strict or $r \prec s_w$).

Generalizing on r we have for each strict or defeasible $r \in R_D[a]$ either $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$, or else r is defeasible and there exists a conflict set $c \in C_D[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R_D[w]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and either $body(r) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $s \not\prec r$ (for ADL, s is strict or $r \prec s$).

Generalizing on a, for each $a \in \mathcal{U}_{E,\lambda}$, for each strict or defeasible $r \in R_D[a]$ either $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$ or else r is defeasible and there exists a conflict set $c \in C_D[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R_D[w]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$ (for ADL, s is strict or $r \prec s$). It can be seen that $\mathcal{U}_{E,\lambda}$ is an unfounded set *wrt* D and $\mathcal{I}_{D,WF}$. And so $\mathcal{U}_{E,\lambda} \subseteq U_D(\mathcal{I}_{D,WF})$. However $\mathcal{U}_{D,WF} = U_D(\mathcal{I}_{D,WF})$. Thus if $q \in \mathcal{U}_{E,\lambda}$, then $q \in \mathcal{U}_{D,WF}$.

If λ is a limit ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{E,\kappa}$ and $p \in \mathcal{T}_{E,\kappa}$. By the inductive hypothesis, $q \in \mathcal{U}_{D,WF}$. □

Theorem 4.45 (Cut for Theories with Transitive priorities)**.** *Let L be NDL or ADL, D a defeasible theory such that $\prec$ is transitive. Let $E = \langle R_D \cup X, C_D, \prec \rangle$ where $X = \{\{\} \rightarrow p\}$ or $X = \{\{\} \Rightarrow p\}$. If $D \mid\!\approx_L p$ and $E \mid\!\approx_L q$, then $D \mid\!\approx_L q$. If $D \mid\!\approx_L p$ and $E \approx\!\mid_L q$, then $D \approx\!\mid_L q$.*

Proof. Since $D \mid\!\approx_L p$, then by cautious monotony, $E \mid\!\approx_L p$. Suppose $E \mid\!\approx_L q$. Then there exists a least $\lambda \geq 0$ such that $p \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{T}_{E,\lambda}$. By Lemma 4.44, $q \in \mathcal{T}_{D,WF}$. Now suppose $E \approx\!\mid_L q$. Then there exists a least $\lambda \geq 0$ such that $p \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{U}_{E,\lambda}$. By Lemma 4.44, $q \in \mathcal{U}_{D,WF}$. □

4.5.3 Rule Simplification

As was stated previously, if the precedence relation is not transitive, then Cut and Cautious Monotony are not satisfied by NDL, and Cautious Monotony is not satisfied by ADL. However, it can be shown that properties somewhat similar to these hold for arbitrarily prioritized theories. Specifically, if $D \mid\!\approx_L p$ and r is a witness of provability for p, then we can safely delete the body of r from the theory without affecting the consequences. That is the best we can do, however. In particular, we must maintain r's place in the precedence relation. This is equivalent to adding a new rule r' with an empty body and the same precedence as r, and it is for this reason that we say that these properties are akin to Cut and Cautious Monotony: we may add a new premise as we do with Cut and Cautious Monotony, but we must ensure that the precedence relation accommodates it in the proper fashion.

We actually make a more general statement (which directly implies that we can delete the body of a witness of provability). Specifically, we claim that a well–founded literal can be deleted from any rule body in which it appears without altering the consequences of a theory. Furthermore, rules containing an unfounded literal in their bodies can be deleted without affecting the consequences of the theory. Theorems for both statements appear below. Both of are useful, for they can be used to simplify further computations. Proofs appear in Appendix C.

Theorem 4.46 (Rule Simplification)**.** *Let L be one of NDL or ADL, D a defeasible theory such that $D \mid\!\approx_L p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) $D \mid\!\approx_L q$ iff $E \mid\!\approx_L q$.
(2) $D \approx\!\mid_L q$ iff $E \approx\!\mid_L q$.

Theorem 4.47 (Rule Elimination)**.** *Let L be one of NDL or ADL, $D = \langle R_D, C_D, \prec_D \rangle$ a defeasible theory such that $D \approx\!\mid_L p$. Let t be any rule such that $p \in body(t)$, and let $E = \langle R_D - \{t\}, C_D, \prec_D \rangle$. For all $q \in Lit_D$,*

(1) $D \mid\!\approx_L q$ iff $E \mid\!\approx_L q$.
(2) $D \approx\!\mid_L q$ iff $E \approx\!\mid_L q$.

4.6 Two Alternating Fixpoint Procedures

In this section and the next we develop operators for so-called alternating fixpoint procedures for computing the results of defeasible theories. The first operator is ambiguity propagating. For unprioritized theories with minimal conflict sets, the consequences obtained using this operator coincide with those of ADL (for technical reasons, the proof of this is postponed until the next chapter). The second operator blocks ambiguity. Its consequences correspond exactly to those of NDL; no restriction on defeasible theories need be made. This is proven in the present chapter.

As with the alternating fixpoint procedure for logic programs, the basic idea is to start with some known set $S \subseteq \mathcal{T}_{WF}$, where $\mathcal{I}_{WF} = \langle \mathcal{T}_{WF}, \mathcal{U}_{WF} \rangle$ is the well-founded model of D. We then determine the reduct of the theory with respect to S and compute its closure. This yields another set of literals S' that is an overestimate of $\mathcal{T}_{WF}$. Since the operation is antimonotone, a double application on S yields a set S'' such that $S \subseteq S'' \subseteq \mathcal{T}_{WF}$. After several such applications, $\mathcal{T}_{WF}$ is obtained.

4.6.1 A Procedure for Propagating Ambiguity

The below operation, since it does not keep track of which rules are used to support a literal, is only defined for defeasible theories in which the precedence relation is empty. Importantly, the procedure can be made to incorporate defeaters: we will allow defeaters to be included in the reduct and their heads to be included in the reduct's closure. However, we will mark such literals

to distinguish them from normal literals. Since no marked literal appears in the body of a rule, no marked literal can be used to support another literal.

Definition 4.48. *Let D be an unprioritized defeasible theory. Then*

$$Lit^*_D = Lit_D \cup \{p^* | p \in Lit_D\}$$

Definition 4.49 (Modified Immediate Consequence)**.** *Let D be an unprioritized defeasible theory and $S \subseteq Lit^*_D$. Then*

$$T^*_D(S) = \{p|\ r \in R_{D,sd}[p] \textit{ and } body(r) \subseteq S\} \cup \{p^*|\ r \in R_{D,u}[p] \textit{ and } body(r) \subseteq S\}$$

We will use $Cl^*(D)$ to mean $T^*_D \uparrow \omega$.

Definition 4.50 (Ambiguity Propagating Reduct for Unprioritized Theories)**.** *Let $D = \langle R, C, \varnothing\rangle$ be an unprioritized defeasible theory and $S \subseteq Lit^*_D$. The reduct of D* wrt *S (written D^S) is the set $R_s \cup R^S_{D,d}$, where*

$$R^S_{D,d} = \{r|\ r \in R_{du} \textit{ and } (\forall c \in C[head(r)])(\exists q \in (c - \{head(r)\}))(q \notin S \textit{ and } q^* \notin S)\}$$

Definition 4.51. *Let D be a defeasible theory and $S \subseteq Lit^*_D$. Then*

$$\alpha_D(S) = T^*_{D^S} \uparrow \omega$$

If D contains no defeaters, then no marked literals will ever appear in $\alpha_D(S)$, regardless of what S contains. When the choice of D is clear from the context, we will drop the subscripts (writing $\alpha(S)$ rather than $\alpha_D(S)$). For finite theories, since the set of rules is finite (and each rule is of finite length), $\alpha(S)$ is finite.

$\langle 2^{Lit^*_D}, \subseteq\rangle$ forms a complete lattice with $\bot = \varnothing$ the bottom element and $\top = Lit^*_D$ the top element. On this lattice, α is antimonotone.

Lemma 4.52 (α is Antimonotone)**.** *If $X \subseteq Y \subseteq Lit^*_D$, then $\alpha_D(Y) \subseteq \alpha_D(X)$.*

Proof. Let R be the rules of D. Since $X \subseteq Y$ every rule deleted from D because of X will be deleted because of Y, $D^Y \subseteq D^X$ and hence $\alpha_D(Y) \subseteq \alpha_D(X)$. □

Definition 4.53. $\delta_D(X) = \alpha_D(\alpha_D(X))$.

Since α is antimonotone, it follows that $\delta(X)$ is monotone. For the complete lattice above, by the Knaster-Tarski theorem a least fixed point exists. Using the above operators, we may define the following sequence.

1) $\delta_D^0(\bot) = \varnothing$

2) $\delta_D^{\lambda+1}(\bot) = \delta_D(\delta_D^{\lambda}(\bot))$ (for successor ordinals)

3) $\delta_D^{\lambda}(\bot) = \bigcup_{\kappa<\lambda} \delta_D^{\kappa}(\bot))$ (for limit ordinals)

When D is finite, δ is continuous on $\langle 2^{Lit_D^*} \subseteq\rangle$, and so $\delta^{\omega}(\varnothing) = lfp(\delta)$. However, as is demonstrated by Example 4.60 in the next section, δ is not in general continuous. From lemma 4.52 and definition of the above sequence, the below proposition immediately follows.

Corollary 4.54. *If $\kappa < \lambda$, then $\alpha(\delta^{\lambda}(\bot)) \subseteq \alpha(\delta^{\kappa}(\bot))$.*

The operators α and δ can be used to define sequences of interpretations.

Definition 4.55. *Let D be an arbitrary unprioritized defeasible theory. Define the sequence of interpretations $\mathcal{I}_0$, $\mathcal{I}_1$, $\mathcal{I}_2$, ... as below.*

(1) $\mathcal{I}_0 = \langle \varnothing, \varnothing \rangle$

(2) $\mathcal{I}_{\lambda} = \langle \delta_D^{\lambda}(\bot), Lit_D - \alpha_D(\delta_D^{\lambda}(\bot)) \rangle$ *(for all $\lambda > 0$).*

The interpretations above may contain too many literals—namely, the marked literals that are the heads of undercutting defeaters. These literals cannot be valid consequences of a theory; they should be discounted. Example 4.44 in the next section shows the sequence of interpretations produced by a theory involving defeaters.

Note that even for finite theories, the iteration of α on the empty set need never reach a single fixed point, but could oscillate between two states X and Y, where $X \subset Y$. The general idea is that the unmarked literals of X constitute the set of well-founded literals *wrt* D. $Lit_D - Y$ is the set of unfounded literals. The unmarked literals in $Y - X$ constitute the set of indeterminate literals, neither well-founded nor unfounded.

The antimonotone operator α described here bears a marked similarity to the GL-operator γ used to define both the stable model/answer set semantics and the WFS. In the next chapter we describe a conceptually simple technique of translating defeasible theories into logic programs. It will be shown there that, under this translation, the two operators coincide.

4.6.2 EXAMPLES

Below are several examples illustrating the alternating fixpoint procedure. Unless stated otherwise, we assume that the precedence relation is empty and conflict sets are closed under strict rules.

Example 4.56. *The well-founded model in this example under both ADL and NDL is* $\langle\{p\}, \{\neg p, q_0, q_1, \ldots\}\rangle$. *The results obtained by iterating* α *from the empty set agree with this.*

$r_1.\ \{\} \Rightarrow p$
$r_2.\ \{q_0\} \Rightarrow \neg p$
$r_3.\ \{q_{n+1}\} \Rightarrow q_n$ (for $n \in \mathbb{N}$)

i	α_D^i	$D^{\alpha_D^i}$	$Cl^*(D^{\alpha_D^i}) = \alpha_D^{i+1}$	$Lit_D - \alpha_D^{i+1}$
0	$\{\}$	$\{r_1, r_2, r_3\}$	$\{p\}$	$\{\neg p, q_0, q_1, \ldots\}$
1	$\{p\}$	$\{r_1, r_3\}$	$\{p\}$	$\{\neg p, q_0, q_1, \ldots\}$
2	$\{p\}$	$\{r_1, r_3\}$	$\{p\}$	$\{\neg p, q_0, q_1, \ldots\}$

Example 4.57. *Here, every literal is undetermined. The well-founded model according to ADL is* $\langle\varnothing, \varnothing\rangle$, *which agrees with the results of* α. *The result under NDL, however, is* $\langle\{\neg r\}, \{p, q, \neg q, r\}\rangle$,

$r_1.\ \{\} \Rightarrow p$
$r_2.\ \{p\} \rightarrow q$
$r_3.\ \{q\} \Rightarrow r$
$r_4.\ \{\} \Rightarrow \neg q$
$r_5.\ \{\} \Rightarrow \neg r$

i	α_D^i	$D^{\alpha_D^i}$	$Cl^*(D^{\alpha_D^i}) = \alpha_D^{i+1}$	$Lit_D - \alpha_D^{i+1}$
0	$\{\}$	$\{r_1, r_2, r_3, r_4, r_5\}$	$\{p, q, r, \neg q, \neg r\}$	$\{\neg p\}$
1	$\{p, q, r, \neg q, \neg r\}$	$\{r_2\}$	$\{\}$	$\{p, \neg p, q, r, \neg q, \neg r\}$
2	$\{\}$	$\{r_1, r_2, r_3, r_4, r_5\}$	$\{p, q, r, \neg q, \neg r\}$	$\{\neg p\}$
3	$\{p, q, r, \neg q, \neg r\}$	$\{r_2\}$	$\{\}$	$\{p, \neg p, q, \neg q, r, \neg r\}$
4	$\{\}$	$\{r_1, r_2, r_3, r_4, r_5\}$	$\{p, q, r, \neg q, \neg r\}$	$\{\neg p\}$

Example 4.58. *Here, every literal has a supporting argument. However, p (a presumption) conflicts with $\neg r$ (a fact), and so p is refutable. Since p is the only support for $\neg t$ and r, both of these are refutable. This leaves t supported with no successful competing arguments, and so t is derivable. The well-founded model according to ADL is $\langle\{t, \neg r\}, \{p, r, \neg t\}\rangle$.*

$r_1.\ \{\} \Rightarrow p$
$r_2.\ \{\} \Rightarrow t$
$r_3.\ \{p\} \rightarrow r$
$r_4.\ \{p\} \Rightarrow \neg t$
$r_5.\ \{\} \rightarrow \neg r$

i	α_D^i	$D^{\alpha_D^i}$	$Cl^*(D^{\alpha_D^i}) = \alpha_D^{i+1}$	$Lit_D - \alpha_D^{i+1}$
0	$\{\}$	$\{r_1, r_2, r_3, r_4, r_5\}$	$\{p, t, r, \neg t, \neg r\}$	$\{\neg p\}$
1	$\{p, t, r, \neg t, \neg r\}$	$\{r_3, r_5\}$	$\{\neg r\}$	$\{p, \neg p, t, r, \neg t\}$
2	$\{\neg r\}$	$\{r_2, r_3, r_4, r_5\}$	$\{t, \neg r\}$	$\{p, \neg p, r, \neg t\}$
3	$\{t, \neg r\}$	$\{r_2, r_3, r_5\}$	$\{t, \neg r\}$	$\{p, \neg p, r, \neg t\}$
4	$\{t, \neg r\}$	$\{r_2, r_3, r_5\}$	$\{t, \neg r\}$	$\{p, \neg p, r, \neg t\}$

Example 4.59. *The below example looks very much like the standard ambiguity test-case. However, it uses a defeater which changes the results. p has no supportive rules and so must be unfounded. This causes q to be unfounded. However, the rule for $\neg p$ (rule 2) conflicts with the supported defeater for p (rule 1). This prevents us from concluding $\neg p$. The well-founded model according to ADL is $\langle\{\neg q\}, \{p, q\}\rangle$. The alternating fixpoint procedure yields the same results. Note that $\neg p$ is ambiguous.*

$r_1.\ \{\} \rightsquigarrow p$
$r_2.\ \{\} \Rightarrow \neg p$
$r_3.\ \{p\} \Rightarrow q$
$r_4.\ \{\} \Rightarrow \neg q$

i	α_D^i	$D^{\alpha_D^i}$	$Cl^*(D^{\alpha_D^i}) = \alpha_D^{i+1}$	$Lit_D - \alpha_D^{i+1}$
0	$\{\}$	$\{r_1, r_2, r_3, r_4\}$	$\{p^*, \neg p, \neg q\}$	$\{p, q\}$
1	$\{p^*, \neg p, \neg q\}$	$\{r_4\}$	$\{\neg q\}$	$\{p, q, \neg p\}$
2	$\{\neg q\}$	$\{r_1, r_2, r_4\}$	$\{p^*, \neg p, \neg q\}$	$\{p, q\}$
3	$\{p^*, \neg p, \neg q\}$	$\{r_4\}$	$\{\neg q\}$	$\{p, q, \neg p\}$
4	$\{\neg q\}$	$\{r_1, r_2, r_4\}$	$\{p^*, \neg p, \neg q\}$	$\{p, q\}$

Example 4.60. *In the below infinite theory, assume that conflict sets are minimal. The theory shows that δ is not continuous. Note that when δ^ω is reached, we have $q_n \in \delta^\omega$ and $\neg p_n \in \delta^\omega$ for all $n \geq 0$. Since this is so, then D^{δ^ω} contains no rule with head $\neg q_n$ for any n. From this it is clear*

that $r \notin \alpha(\delta^{\omega}) = Cl^{*}(D^{\delta^{\omega}})$. *Furthermore, we have* $\neg r \in \alpha(\delta^{\omega})$, *and for all* n, $q_n \in \alpha(\delta^{\omega})$ *and* $\neg p_n \in \alpha(\delta^{\omega})$. *Applying* α *again yields the same results. Thus a fixpoint is reached at* $\alpha(\alpha(\delta^{\omega})) = \delta^{\omega+1}$ *and not before.*

$r_1.\ \{\neg p_n\} \rightarrow q_n$ $r_2.\ \{\neg q_n\} \rightarrow p_{n+1}$ $r_3.\ \{\} \Rightarrow \neg p_n$ $r_4.\ \{\} \Rightarrow \neg q_n$ $r_5.\ \{\neg q_n\} \Rightarrow r$ $r_6.\ \{\} \Rightarrow \neg r$ *(for all* $n \in \mathbb{N}$*)*	λ	$\alpha_D^{\lambda}(\bot)$
	1	$\{p_1, p_2, \ldots, q_0, q_1, \ldots, \neg p_0, \neg p_1, \ldots, \neg q_0, \neg q_1, \ldots, r, \neg r\}$
	2	$\{q_0, \neg p_0\}$
	3	$\{p_2, p_3, \ldots, q_0, q_1, \ldots, \neg p_0, \neg p_1, \ldots, \neg q_1, \neg q_2, \ldots, r, \neg r\}$
	4	$\{q_0, q_1, \neg p_0, \neg p_1\}$
	5	$\{p_3, p_4, \ldots, q_0, q_1, \ldots, \neg p_0, \neg p_1, \ldots, \neg q_2, \neg q_3, \ldots, r, \neg r\}$
	6	$\{q_0, q_1, q_2, \neg p_0, \neg p_1, \neg p_2\}$
	7	etc.

4.6.3 THE PROCEDURE FOR NDL

The above fixpoint procedure obviously is ambiguity propagating. However, it can be modified to yield an ambiguity blocking relation. Indeed, we show here that the modified alternating fixpoint procedure yields the well-founded model according to NDL.

Definition 4.61 (Ambiguity Blocking Reduct for Unprioritized Theories)**.** *Let* $D = \langle R, C, \prec \rangle$ *be a defeasible theory and* $S \subseteq Lit_D$. *The blocking–reduct of* D wrt S *(written* D_{β}^{S}*) is* $R_{D,s} \cup R_{D,d}^{S}$ *where*

$$R_{D,d}^{S} = \{r |\ r \in R_d[p] \textit{ and } (\forall c \in C[p])(\exists q \in (c - \{p\}))(\forall s \in R[q])(body(s) \nsubseteq S \textit{ or } s \prec r\}$$

Unlike the ambiguity propagating counterpart, we do not include the heads of defeaters in the closure of the reduct, nor for that matter are defeaters included in the reduct. In the definition, we look only at the bodies of rules in computing the reduct and not the heads.

Definition 4.62. *Let* D *be a defeasible theory and* $S \subseteq Lit_D$. *Define* $\beta_D(S)$ *to be* $Cl(D_{\beta}^{S})$.

All of the lemmas stated for α have counterparts that hold for β. In the following we will use δ_β to stand for the ambiguity counterpart of δ. We will also use β^n to stand for $\beta^n(\varnothing)$ and δ^n_β to stand for $\delta^n_\beta(\varnothing)$, respectively.

Theorem 4.63. *Let D be a defeasible theory and $\mathcal{I}_{D,WF} = \langle \mathcal{T}_{D,WF}, \mathcal{U}_{D,WF} \rangle$ the well-founded model of D* wrt *NDL. For all $\lambda \geq 0$, if $p \in \delta^\lambda_\beta$, then $p \in \mathcal{T}_{D,WF}$. If $p \notin \beta(\delta^\lambda_\beta)$, then $p \in \mathcal{U}_{D,WF}$.*

Proof. For $\lambda = 0$, $\delta^0_\beta = \varnothing$ and so no $p \in \delta^0_\beta$. If $p \notin \beta(\delta^0_\beta)$ then obviously it is impossible to derive p from the rules of D under any circumstances (p is unreachable relative to D's entire set of rules), and so p is in $\mathcal{U}_{D,WF}$. Suppose the hypothesis holds for all $\kappa < \lambda$.

(Case 1) Suppose $p \in \delta^n_\beta$. Then there exists a least successor ordinal $\kappa \leq \lambda$ such that $p \in \delta^\kappa_\beta$. Recall the following.

$$\delta^\kappa_\beta = \beta(\beta(\delta^{\kappa-1}_\beta)) = Cl(D_\beta^{\beta(\delta^{\kappa-1}_\beta)}) = T_{D_\beta^{\beta(\delta^{\kappa-1}_\beta)}} \uparrow \omega$$

To simplify the notation, we will use the below convention in the remainder of the proof.

$$T[D_\beta^{\beta(\delta^{\kappa-1}_\beta)}] \uparrow \omega \text{ stands for } T_{D_\beta^{\beta(\delta^{\kappa-1}_\beta)}} \uparrow \omega.$$

Suppose that for all $i < m$, if $a \in T[D_\beta^{\beta(\delta^{\kappa-1}_\beta)}] \uparrow i$, then $a \in \mathcal{T}_{D,WF}$ (this obviously holds for $i = 0$). Let $p \in T[D_\beta^{\beta(\delta^{\kappa-1}_\beta)}] \uparrow m$. Then there is an r in $D_\beta^{\beta(\delta^{\kappa-1}_\beta)}$ such that $body(r) \subseteq T[D_\beta^{\beta(\delta^{\kappa-1}_\beta)}] \uparrow (m-1)$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D,WF}$.

If r is strict, then obviously $p \in \mathcal{T}_{D,WF}$ by definition of T_D and $\mathcal{I}_{D,WF}$. If r is defeasible, then since $r \in D_\beta^{\beta(\delta^{\kappa-1}_\beta)}$, for all conflict sets $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for each rule $s \in R[q]$, either $s \prec r$ or $body(s) \not\subseteq \beta(\delta^{\kappa-1}_\beta)$. If the latter, then by the inductive hypothesis, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Then there is a rule r of D such that r is strict and $body(r) \subseteq \mathcal{T}_{D,WF}$, or else r is defeasible, $body(r) \subseteq \mathcal{T}_{D,WF}$, and for all conflict sets $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for each rule $s \in R[q]$, either $s \prec r$ or $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Thus, by definition of immediate consequence in NDL and $\mathcal{I}_{D,WF}$, $p \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose $p \notin \beta(\delta^\lambda_\beta)$, and let X be the set of elements not in $\beta(\delta^\lambda_\beta)$. Let r be a rule for p. If r is strict, then r is in the reduct of D relative to δ^λ_β and there is some $q \in body(r)$ such that

$q \in X$ (this must be the case since $\beta(\delta_\beta^\lambda)$ is closed). If r is defeasible then r is either in the reduct or not. If it is, then as before there is some $q \in body(r)$ such that $q \in X$. If not, then there is a conflict set $c \in C[p]$ such that for all $q \in c - \{p\}$, there is a rule $s \in R[q]$ such that $body(s) \subseteq \delta_\beta^\lambda$ and $s \not\prec r$. From Case 1, if $body(s) \subseteq \delta_\beta^\lambda$ then $body(s) \subseteq \mathcal{T}_{D,WF}$. Generalizing on r and then on p, for each $a \in X$ and each $r \in R_{D,sd}[a]$, either there exists a $q \in body(r)$ such that $q \in X$ or else $r \in R_d[a]$ and there exists a conflict set $c \in C[a]$ such that for each $v \in c - \{a\}$ there is a rule $s \in R[v]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$. It can be seen that X is unfounded under NDL with respect to D and $\mathcal{I}_{D,WF}$. As such $X \subseteq U_D(\mathcal{U}_{D,WF})$, and in particular, $p \in U_D(\mathcal{U}_{D,WF})$. Since $U_D(\mathcal{U}_{D,WF}) = \mathcal{U}_{D,WF}$, $p \in \mathcal{U}_{D,WF}$. □

Theorem 4.64. *If D is a defeasible theory and $(\mathcal{I}_D)$ the sequence of interpretations defined for D under the NDL-well-founded semantics, then for any $\lambda \geq 0$, if $p \in \mathcal{T}_{D,\lambda}$, then there exists a $\eta \geq 0$ such that $p \in \delta_\beta^\eta$. If $p \in \mathcal{U}_{D,\lambda}$, then there exists an $\eta \geq 0$ such that $p \notin \beta(\delta_\beta^\eta)$.*

Proof. The hypothesis is trivially satisfied for $\lambda = 0$. Suppose the hypothesis holds for all $\kappa < \lambda$.
(Case 1): Let $p \in \mathcal{T}_\lambda$. Then one of two cases applies: (1.1) there is a successful strict rule or (1.2) there is a successful defeasible rule. We consider each in turn.
(1.1) There exists a rule $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_\kappa$ for some successor ordinal $\kappa < \lambda$. If that is the case, then by the inductive hypothesis, there exists an $\eta \geq 0$ such that $body(r) \subseteq \delta_\beta^\eta$. Since r is strict (and thus appears in all reducts) and δ_β^η is closed under strict rules, $p \in \delta_\beta^\eta$.
(1.2) there exists a defeasible rule r such that $body(r) \subseteq \mathcal{T}_\kappa$ for some $\kappa < \lambda$ and for all $c \in C[p]$ there is a $q \in c - \{p\}$ such that for all rules $s \in R[q]$, either $s \prec r$ or else there exists a $v \in body(s)$ such that $v \in \mathcal{U}_\kappa$. If the latter, then by the inductive hypothesis, there exists a η such that $v \notin \beta(\delta_\beta^\eta)$. Since $body(r) \subseteq \mathcal{T}_\kappa$, then by the inductive hypothesis, $body(r) \subseteq \delta_{\beta^\iota}$ for some ι. Note that for any ordinals α and γ, if $\alpha < \gamma$, then $\delta_\beta^\alpha \subseteq \delta_\beta^\gamma$ and $\beta(\delta_\beta^\gamma) \subseteq \beta(\delta_\beta^\alpha)$, and so for any literal b, if $b \notin \beta(\delta_\beta^\alpha)$, then for all $\gamma > \alpha$ it follows that $b \notin \beta(\delta_\beta^\gamma)$. With that in mind, generalizing on s and then c, and letting μ be the least ordinal such that $\eta < \mu$ and $\iota < \mu$ for any of the above ι's, we have $body(r) \subseteq \delta_\beta^\mu$ and for all $c \in C[p]$ there is a $q \in c - \{p\}$ such that for each $s \in R[q]$, $body(s) \nsubseteq \beta(\delta_\beta^\mu)$ or $s \prec r$. As such $r \in D^{\beta(\delta_\beta^\mu)}$ by definition of blocking–reducts. Since $body(r) \subseteq \delta_\beta^\mu$, by monotonicity

we have $body(r) \subseteq \delta_\beta^{\mu+1}$. Recall that $\delta_\beta^{\mu+1} = \beta(\beta(\delta_\beta^\mu)) = Cl(D^{\beta(\delta_\beta^\mu)})$. We thus have $r \in D^{\beta(\delta_\beta^\mu)}$ and $body(r) \subseteq Cl(D^{\beta(\delta_\beta^\mu)})$. From this it follows that $p \in Cl(D^{\beta(\delta_\beta^\mu)})$ (*i.e.*, $p \in \delta_\beta^{\mu+1}$).

(Case 2): Suppose $p \in \mathcal{U}_\lambda$.

(Case 2.1): If λ is a limit ordinal, then there exists some successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_\kappa$. By the inductive hypothesis, there exists some $\eta \geq 0$ such that $p \notin \beta(\delta_\beta^\eta)$.

(Case 2.2): Suppose λ is a successor ordinal. By definition, $\mathcal{U}_\lambda$ is an unfounded set *wrt* to D and $\mathcal{I}_{\lambda-1}$. As this is so, for all $a \in \mathcal{U}_\lambda$, for each rule $r \in R_{s,d}[a]$, either (2.2.1) there is a $v \in body(r)$ such that $v \in \mathcal{U}_\lambda \cup \mathcal{U}_{\lambda-1}$ (which by monotonicity of U_D means $v \in \mathcal{U}_\lambda$) or else (2.2.2) $r \in R_d[q]$ and there exists a conflict set $c \in C[a]$ such that for each $q \in c - \{a\}$, there is a rule $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_{\lambda-1}$ and $s \not\prec r$. Suppose (2.2.2) holds. Since $body(s) \subseteq \mathcal{T}_{\lambda-1}$, then by the inductive hypothesis, there exists a $\gamma \geq 0$ such that $body(s) \subseteq \delta_\beta^\gamma$. Generalizing on q, there exists a $\eta \geq 0$ such that for each $q \in c - \{a\}$ there exists a $s \in R[a]$ such that $body(s) \subseteq \delta_\beta^\eta$ and $s \not\prec r$. As this is so, by definition of reduct for NDL, it must be the case that $r \notin D_\beta^{\delta_\beta^\eta}$ (i.e, r is not in the reduct of D relative to δ_β^η).

Thus, for all rules $r \in R[a]$, if (2.2.2) holds, then $r \notin D_\beta^{\delta_\beta^\eta}$. If $r \in D_\beta^{\delta_\beta^\eta}$, then (2.2.1) must hold. Generalizing on a, for each $v \in \mathcal{U}_\lambda$, if $r \in R_{sd}[v]$ and $r \in D_\beta^{\delta_\beta^\eta}$, then $body(r) \cap \mathcal{U}_\lambda \neq \varnothing$. As this is so, $\mathcal{U}_\lambda \cap Cl(D_\beta^{\delta_\beta^\eta}) = \varnothing$.[1] That is, no element of $\mathcal{U}_\lambda$ is an element of $\beta(\delta_\beta^\eta)$. In particular, $p \notin \beta(\delta_\beta^\eta)$. □

From the above theorems, the correspondence between the alternating fixpoint procedure and the well-founded model according to NDL is established.

Corollary 4.65. *Let D be a defeasible theory and $\mathcal{I}_{D,WF}$ its well-founded model according to NDL. Then $\mathcal{I}_{D,WF} = \langle \delta_\beta^\infty(\varnothing), Lit_D - \beta(\delta_\beta^\infty(\varnothing))\rangle$.*

[1] There is no way to get the derivation started: There must be a least $i > 0$ such that $\mathcal{U}_\lambda \cap T[D_\beta^{\delta_\beta^\eta}] \uparrow i \neq \varnothing$. However, if $v \in T[D_\beta^{\delta_\beta^\eta}] \uparrow i$, then there exists a rule $r \in R_{sd}[v]$ such that $body(r) \subseteq T[D_\beta^{\delta_\beta^\eta}] \uparrow (i-1)$. Thus there is some $q \in (\mathcal{U}_\lambda \cap T[D_\beta^{\delta_\beta^\eta}] \uparrow (i-1))$. A contradiction.

4.6.4 EXAMPLES REVISITED

We return to a few of the examples from previous sections to illustrate the behavior of the ambiguity blocking operator. As before, assume that the precedence relation is empty. Also assume conflict sets are closed under strict rules.

Example 4.66. *Here, p and $\neg q$ conflict. Since neither rule 1 or 4 is preferred, both literals are unfounded in NDL. Thus both q and r are undercut, leaving one free to conclude $\neg r$. The well-founded model according to NDL is $\langle\{\neg r\}, \{p, \neg p, q, \neg q, r\}\rangle$. This agrees with the results of β.*

$r_1.\ \{\} \Rightarrow p$
$r_2.\ \{p\} \rightarrow q$
$r_3.\ \{q\} \Rightarrow r$
$r_4.\ \{\} \Rightarrow \neg q$
$r_5.\ \{\} \Rightarrow \neg r$

i	α_D^i	$D^{\alpha_D^i}$	$Cl(D^{\alpha_D^i}) = \alpha_D^{i+1}$	$Lit_D - \alpha_D^{i+1}$
0	$\{\}$	$\{r_2, r_5\}$	$\{\neg r\}$	$\{p, q, r, \neg q\}$
1	$\{\neg r\}$	$\{r_2, r_5\}$	$\{\neg r\}$	$\{p, q, r, \neg p, \neg q\}$
2	$\{\neg r\}$	$\{r_2, r_5\}$	$\{\neg r\}$	$\{p, q, r, \neg p, \neg q\}$
3	$\{\neg r\}$	$\{r_2, r_5\}$	$\{\neg r\}$	$\{p, q, r, \neg p, \neg q\}$
4	$\{\neg r\}$	$\{r_2, r_5\}$	$\{\neg r\}$	$\{p, q, r, \neg p, \neg q\}$

Example 4.67. *Below is the standard ambiguity test-case. Rule 1 conflicts with rule 2. Since neither is preferred and both are supported, neither is included in the reduct. Rule 4 conflicts with rule 3. However, rule 3 is not supported initially and so does not prevent rule 4 from entering the reduct. Rule 4 is supported, and this prevents inclusion of rule 3. The well-founded model according to NDL is $\langle\{\neg q\}, \{p, q, \neg p\}\rangle$. This agrees with the results of β.*

$r_1.\ \{\} \Rightarrow p$
$r_2.\ \{\} \Rightarrow \neg p$
$r_3.\ \{p\} \Rightarrow q$
$r_4.\ \{\} \Rightarrow \neg q$

i	α_D^i	$D^{\alpha_D^i}$	$Cl(D^{\alpha_D^i}) = \alpha_D^{i+1}$	$Lit_D - \alpha_D^{i+1}$
0	$\{\}$	$\{r_4\}$	$\{\neg q\}$	$\{p, q, \neg p\}$
1	$\{\neg q\}$	$\{r_4\}$	$\{\neg q\}$	$\{p, q, \neg p\}$
2	$\{\neg q\}$	$\{r_4\}$	$\{\neg q\}$	$\{p, q, \neg p\}$

Example 4.68. *(Example 4.60) This example also shows that δ_β is not continuous in general. The only real distinction here is that since rule 6 is always (vacuously) supported, no rule with head r ever appears in any reduct. As this is so, r itself never appears in the closure of any reduct.*

$r_1.\ \{\neg p_n\} \rightarrow q_n$ $r_2.\ \{\neg q_n\} \rightarrow p_{n+1}$ $r_3.\ \{\} \Rightarrow \neg p_n$ $r_4.\ \{\} \Rightarrow \neg q_n$ $r_5.\ \{\neg q_n\} \Rightarrow r$ $r_6.\ \{\} \Rightarrow \neg r$ *(for all $n \in \mathbb{N}$)*	λ	$\alpha_D^\lambda(\bot)$
	1	$\{p_1, p_2, \ldots, q_0, q_1, \ldots, \neg p_0, \neg p_1, \ldots, \neg q_0, \neg q_1, \ldots, \neg r\}$
	2	$\{q_0, \neg p_0\}$
	3	$\{p_2, p_3, \ldots, q_0, q_1, \ldots, \neg p_0, \neg p_1, \ldots, \neg q_1, \neg q_2, \ldots, \neg r\}$
	4	$\{q_0, q_1, \neg p_0, \neg p_1\}$
	5	$\{p_3, p_4, \ldots, q_0, q_1, \ldots, \neg p_0, \neg p_1, \ldots, \neg q_2, \neg q_3, \ldots, \neg r\}$
	6	$\{q_0, q_1, q_2, \neg p_0, \neg p_1, \neg p_2\}$
	7	etc.

CHAPTER 5

RELATING DEFEASIBLE LOGIC TO LOGIC PROGRAMMING

For unprioritized defeasible theories with minimal conflict sets and no defeaters, the semantics for ADL corresponds to the well-founded semantics for logic programs. Specifically, there exists a rather natural translation of defeasible theories into logic programs which preserves the results of the theories. Furthermore, there is a separate translation of logic programs into defeasible theories which preserves the results of the programs. Thus, either formalism can be embedded in the other. This is useful, for it means that logic programming systems for the WFS can be used to compute the consequences of ADL. Also, under the same translation scheme, the semantics for NDL can be viewed as providing an ambiguity blocking semantics for logic programs.

The translation of defeasible theories into logic programs does not in fact require minimal conflict sets. The correspondence with ADL only holds if this requirement is met, however. Later sections of the present chapter show that under the lesser restriction that $C[p]$ is finite for all p, the antimonotone operator α defined previously corresponds to the GL-operator γ. Since ADL corresponds to the WFS when conflict sets are minimal, it follows that when conflict sets are minimal a correspondence holds between the alternating fixpoint procedure based on α and the consequences of ADL.

5.1 TRANSLATING DEFEASIBLE THEORIES INTO LOGIC PROGRAMS

Brewka [Bre01] provides a simple translation scheme to compare a version of defeasible logic (specifically, BDLA without team defeat [AM02]) to logic programs using his own prioritized well-founded semantics. Several examples are presented to demonstrate that the two systems do not agree, and Brewka argues that the results of the defeasible theory are less than reasonable.

Brewka dismisses without examination other versions of defeasible logic because they are ambiguity blocking.

We have here altered the translation scheme to encompass theories with extended conflict sets; we will use it to compare ADL to the simple WFS. Since conflict sets are sufficient to encode negation, we will assume all literals appearing in a defeasible theory are positive ($\neg p$ is treated as an atom unrelated to p). Furthermore, since defeaters and the precedence relation are not defined for the simple WFS, we will assume that no defeaters occur in the theory and that the precedence relation is empty.

Definition 5.1. *Let $D = \langle R, C, \prec \rangle$ be a defeasible theory such that $C[p] = (c_1, c_2, \ldots, c_m)$. For any literal $p \in Lit_D$,*

$$Prod(C[p]) = \{\{a_1, a_2, \ldots, a_m\} | \{a_1, a_2, \ldots, a_m\} \in \prod_{i=1}^{m} (c_i - \{p\})\}$$

$Prod(C[p])$ is the set of all sets that can be created by taking a single literal (other than p) from each conflict set containing p. We use these sets when translating defeasible rules of a theory into logic program rules. For the sake of the translation, we require that $C[p]$ be finite for each $p \in Lit_D$ and that each $c \in C[p]$ be finite. This entails that each set in $Prod(C[p])$ is finite. This is needed to ensure that the resulting rules of the logic program are finite in length (infinite rules are usually not allowed). The requirement is obviously met if D is finite.

Definition 5.2 (Logic Program Translation)**.** *Let $D = \langle R, C, \varnothing \rangle$ be a defeasible theory such that $R_u = \varnothing$ and for each $p \in Lit_D$, $C[p]$ is finite. The logic program translation of D, denoted Π_D, is the smallest set of rules such that*

(1) If $\{q_1, q_2, \ldots, q_n\} \rightarrow p \in R_s$, then p :- $q_1, q_2, \ldots, q_n \in \Pi_D$.
(2) If $\{q_1, q_2, \ldots, q_n\} \Rightarrow p \in R_d$ and $\{a_1, \ldots a_m\} \in Prod(C[p])$, then

p :- *not* $a_1, \ldots,$ *not* $a_m, q_1, q_2, \ldots, q_n \in \Pi_D$.

Let $trans(r)$ denote the set of logic program rules obtained from rule r of a defeasible theory. There is a 1-1 correspondence between strict rules of D and rules of Π_D. That is, when r is strict, $trans(r) = \{r\}$ (we may ignore the notational differences; we also choose to ignore the braces,

saying instead that $trans(r) = r$). For each defeasible rule r of D, however, there may be many rules of Π_D. Each a_i is a literal of some conflict set containing p. For r to succeed, at least one literal from each conflict set must be unfounded. Significantly, if conflict sets are minimal, then a 1-1 correspondence between defeasible rules and rules of Π_D again obtains. That is, if $r = \{q_1, q_2 \ldots q_n\} \Rightarrow p$, then $trans(r) = \{p \ :- \ not \ \neg p, q_1, q_2, \ldots, q_n\}$. We observe that Brewka only considers theories where conflict sets are minimal.

Returning to Example 1.3, the logic program translation based on extended conflict sets is shown below.

Example 5.3. *(Ex 1.3)*

(1) $livesAlone(joe)$.
(2) $bachelor(joe) \ :- \ not \ \neg bachelor(joe), \ not \ married(joe), livesAlone(joe)$.
(3) $\neg married(joe) \ :- \ not \ married(joe), bachelor(joe)$.
(4) $married(joe)$.

With minimal conflict sets, the translation is

(1) $livesAlone(joe)$.
(2) $bachelor(joe) \ :- \ not \ \neg bachelor(joe), livesAlone(joe)$.
(3) $\neg married(joe) \ :- \ not \ married(joe), bachelor(joe)$.
(4) $married(joe)$.

In the first case, in order to show that Joe is a bachelor, we must first show that we cannot derive $\neg bachelor$ and $married$. In the WFS for logic programs, $bachelor(joe)$ is unfounded, and this is precisely the result we want. In the second case, since there are no rules for $\neg bachelor(joe)$, this literal is unfounded, and so both $bachelor(joe)$ and $married(joe)$ are well-founded. This example shows that we must consider extended conflict sets in the translation in order to achieve the correct results.

5.2 Soundness and Completeness

Under restrictions, ADL is sound and complete *wrt* the WFS. We prove these claims now. Since the operators for the WFS have direct analogs for defeasible logic and are defined individually for

each logic program/defeasible theory, we can use the same basic symbols for each (writing, for instance, T_D and T_Π) without causing confusion. We will use $\mathcal{I}_{D,0}, \mathcal{I}_{D,1}, \ldots$ to denote the sequence of interpretations obtained using the operator W_D, and $\mathcal{I}_{\Pi,0}, \mathcal{I}_{\Pi,1}, \ldots$ to denote the sequence of interpretations obtained by the operator W_Π. For a given $\mathcal{I}_{\Pi,\lambda}$, we will write $\mathcal{T}_{\Pi,\lambda}$ and $\mathcal{U}_{\Pi,\lambda}$ to distinguish well-founded and unfounded sets in logic programs from their defeasible logic counterparts $\mathcal{T}_{D,\lambda}$ and $\mathcal{U}_{D,\lambda}$.

5.2.1 SOUNDNESS

Theorem 5.4 (Soundness under the translation)**.** *Let $D = \langle R, C, \varnothing\rangle$ be a defeasible theory such that $R_u = \varnothing$ and for each $p \in Lit_D$, $C[p]$ is finite. Let Π be the logic program translation of D. For any $p \in Lit_D$, if $D \mid\!\!\approx_{ADL} p$, then $\Pi \mid\!\!\approx_{WFS} p$, and if $D \approx\!\!|_{ADL} p$, then $\Pi \approx\!\!|_{WFS} p$.*

Proof. The proof is by induction on the sequence $\mathcal{I}_{D,0}, \mathcal{I}_{D,1}, \ldots$. We will show that for all $\lambda \geq 0$, if $p \in \mathcal{T}_{D,\lambda}$, then $p \in \mathcal{T}_{\Pi,WF}$, and if $p \in \mathcal{U}_{D,\lambda}$, then $p \in \mathcal{U}_{\Pi,WF}$. Since $\mathcal{I}_{D,0} = \mathcal{I}_{\Pi,0}$, the hypothesis holds for $\lambda = 0$. Suppose that the hypothesis holds for all $\kappa < \lambda$. We proceed by cases.

(Case 1) Let $p \in \mathcal{T}_{D,\lambda}$ and suppose that λ is a successor ordinal. Then there exists an $r \in R[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$ and either (1) r is strict or else (2) r is defeasible and for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R[q]$, $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. In both cases, $body(r) \subseteq \mathcal{T}_{\Pi,WF}$ by the inductive hypothesis. As such, for any $r' \in trans(r)$, $body(r')^+ \subseteq \mathcal{T}_{\Pi,WF}$. Furthermore, if r is strict (*i.e.*, (1) obtains) then $body(r')^+ = body(r')$ and so $body(r') \subseteq \mathcal{T}_{\Pi,WF}$. By definition of T_Π, $p \in \mathcal{T}_{\Pi,WF}$.

Suppose (2) obtains. Then for all $t \in trans(r)$, $body(t)^+ \subseteq \mathcal{T}_{\Pi,WF}$. Let $c \in C[p]$. Then there exists a $q \in c - \{p\}$ such that for each $s \in R[q]$, $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. Let $s' \in trans(s)$. Since $body(s) = body(s')^+$, $body(s')^+ \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. By inductive hypothesis, $body(s')^+ \cap \mathcal{U}_{\Pi,WF} \neq \varnothing$. Since every logic program rule for q has a classical literal in $\mathcal{U}_{\Pi,WF}$, it follows by definition of U_Π and $\mathcal{I}_{\Pi,WF}$ that $q \in \mathcal{U}_{\Pi,WF}$.

Generalizing on q, every conflict set for p has a literal $q_i \neq p$ such that $q_i \in \mathcal{U}_{\Pi,WF}$. Suppose $|C[p]| = m$ and let $Q = \{q_1, q_2, \ldots, q_m\}$ be a set of such q's. Obviously, $Q \subseteq Prod(C[p])$ and

so there is a particular rule $r' \in trans(r)$ such that $body(r')^+ = body(r)$ and $body(r')^- = Q$. Given the above discussion, $body(r')^+ \subseteq \mathcal{T}_{\Pi,WF}$ and $body(r')^- \subseteq \mathcal{U}_{\Pi,WF}$. By definition of T_Π and $\mathcal{I}_{\Pi,WF}$, $p \in \mathcal{T}_{\Pi,WF}$.

If λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{D,\kappa}$. By the inductive hypothesis, $p \in \mathcal{T}_{\Pi,WF}$.

(Case 2) Let $p \in \mathcal{U}_{D,\lambda}$ and suppose that λ is a successor ordinal. Note that $\mathcal{U}_{D,\lambda}$ is by definition unfounded *wrt* D and $\mathcal{I}_{D,\lambda-1}$. Suppose $r \in R_s[p]$. Then there is a $q \in body(r)$ such that $q \in \mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}$. Since U_D is monotone, $q \in \mathcal{U}_{D,\lambda}$.

Suppose $r \in R_d[p]$. Then either (1) there is a $q \in body(r)$ such that $q \in \mathcal{U}_{D,\lambda}$, or (2) there is a conflict set $c \in C[p]$ such that for all $a \in c - \{p\}$, there is a $s \in R_s[a]$ such that $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ (s must be strict since the priority relation is empty). Suppose (2) holds. By inductive hypothesis, $body(s) \subseteq \mathcal{T}_{\Pi,WF}$. Since s is strict, $trans(s) = s$, and so $a \in \mathcal{T}_{\Pi,WF}$ by definition of T_Π and $\mathcal{I}_{\Pi,WF}$. Recall that by definition of $Prod(C[p])$, for each set $Q \in Prod(C[p])$ we have $Q \cap (c - \{p\}) \neq \varnothing$. By definition of $trans(r)$, for each $t \in trans(r)$, there exists a $Q \in Prod(C[p])$ such that $Q = body(t)^-$. Since this is so, if (2) holds then for each rule $r' \in trans(r)$, there exists a $not\ a \in body(r')^-$ such that $a \in \mathcal{T}_{\Pi,WF}$.

Generalizing on r, every logic program rule r' for p has a classical literal $q \in body(r')^+$ such that $q \in \mathcal{U}_{D,\lambda}$, or else a default literal $a \in body(r')^-$ such that $a \in \mathcal{T}_{\Pi,WF}$. Generalizing on p, for each $b \in \mathcal{U}_{D,\lambda}$ and for each rule t with head b, either there exists a classical subgoal $q \in body(t)^+$ such that $q \in \mathcal{U}_{D,\lambda}$ or else a default literal $a \in body(t)^-$ such that $a \in \mathcal{T}_{\Pi,WF}$. Thus by definition of unfounded sets for logic programs, $\mathcal{U}_{D,\lambda}$ is an unfounded set relative to Π and $\mathcal{I}_{\Pi,WF}$, and so $\mathcal{U}_{D,\lambda} \subseteq \mathcal{U}_{\Pi,WF}$. Since $p \in \mathcal{U}_{D,\lambda}$, it follows that $p \in \mathcal{U}_{\Pi,WF}$.

If λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{D,\kappa}$. By the inductive hypothesis, $p \in \mathcal{U}_{\Pi,WF}$. □

Note that the claim of soundness encompasses defeasible theories with extended conflict sets, provided that $C[p]$ is finite for all p. The other specific requirements are that no defeaters are used

and that the rules are unprioritized. A similar claim cannot be made for completeness, however. Completeness *does* require conflict sets to be minimal.

Before we proceed, the following lemma is needed.

Lemma 5.5. *Let $D = \langle R, C, \varnothing\rangle$ be a defeasible theory such that $R_u = \varnothing$ and each $c \in C$ is minimal. If $r \in R_d[p]$ and $body(r) \subseteq \mathcal{T}_{WF}$, and $\neg p \in \mathcal{U}_{WF}$, then $p \in \mathcal{T}_{WF}$.*

Proof. Suppose $r \in R_d[p]$ and $body(r) \in \mathcal{T}_{WF}$ and $\neg p \in \mathcal{U}_{WF}$. Then there must be some least successor ordinal λ such that $body(r) \in \mathcal{T}_\lambda$ and $\neg p \in \mathcal{U}_\lambda$. Recall that $\mathcal{U}_\lambda$ is the greatest unfounded set *wrt* $\mathcal{I}_{\lambda-1}$. By definition of unfounded set (and since conflict sets are minimal) we have,

(1) for all $s \in R_s[\neg p]$, $body(s) \cap (\mathcal{U}_\lambda \cup \mathcal{U}_{\lambda-1}) \neq \varnothing$, and
(2) for all $s \in R_d[\neg p]$, either

  (2.1) $body(s) \cap (\mathcal{U}_\lambda \cup \mathcal{U}_{\lambda-1}) \neq \varnothing$, or
  (2.2) there is a rule $t \in R_s[p]$ such that $body(t) \subseteq \mathcal{T}_{\lambda-1}$.

If 2.2 above holds, then $p \in \mathcal{T}_\lambda$. Similarly, since U_D is monotone, $body(s) \cap (\mathcal{U}_\lambda \cup \mathcal{U}_{\lambda-1})$ can be reduced to $body(s) \cap \mathcal{U}_\lambda$. Thus we have $r \in R_d[p]$ and $body(r) \subseteq \mathcal{T}_\lambda$, and for each rule $s \in R[\neg p]$, either $body(s) \cap \mathcal{U}_\lambda \neq \varnothing$ or $p \in \mathcal{T}_\lambda$. Since the second disjunct does not involve s, we may write[1]: Either $p \in \mathcal{T}_\lambda$ or for all rules $s \in R[\neg p]$, $body(s) \cap \mathcal{U}_\lambda \neq \varnothing$. In the first case, $p \in \mathcal{T}_{\lambda+1}$ by monotonicity of the sequence $(\mathcal{I})$. In the second case $p \in \mathcal{T}_{\lambda+1}$ by the immediate consequence operator. □

5.2.2 COMPLETENESS

Theorem 5.6 (Completeness under the translation)**.** *Let $D = \langle R, C, \varnothing\rangle$ be a defeasible theory such that $R_u = \varnothing$ and each $c \in C$ is minimal, and let Π be its logic program translation. Then for all $p \in Lit_D$, if $\Pi \mid\!\approx_{WFS} p$, then $D \mid\!\approx_{ADL} p$, and if $\Pi \not\approx_{WFS} p$, then $D \not\approx_{ADL} p$.*

Proof. Let $\mathcal{I}_{D,WF} = \langle \mathcal{T}_{D,WF}, \mathcal{U}_{D,WF}\rangle$ be the well–founded model of D under ADL. The proof is by induction on the sequence $\mathcal{I}_{\Pi,0}$, $\mathcal{I}_{\Pi,1}$, ... Suppose that for all $\kappa < \lambda$, if $p \in \mathcal{T}_{\Pi,\kappa}$, then $p \in \mathcal{T}_{D,WF}$, and if $p \in \mathcal{U}_{\Pi,\kappa}$, then $p \in \mathcal{U}_{D,WF}$. Since $\mathcal{I}_{\Pi,0} = \mathcal{I}_{D,0}$, the hypothesis holds for $\kappa = 0$.

[1] This is an instance of $\forall(x)(q(x) \vee P) \rightarrow \forall(x)q(x) \vee P$, where P contains no free occurrence of x.

(Case 1) Suppose $p \in \mathcal{T}_{\Pi,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $s \in trans(r)$ for some $r \in R[p]$ such that $body(s)^+ \subseteq \mathcal{T}_{\Pi,\lambda-1}$ and $body(s)^- \subseteq \mathcal{U}_{\Pi,\lambda-1}$. $body(s)^+ = body(r)$ and so by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$. If $body(s)^- = \varnothing$, then $r \in R_s[p]$ and $p \in \mathcal{T}_{D,WF}$ by definition of $\mathcal{I}_{D,WF}$ and T_D for defeasible theories. If $body(s)^- \neq \varnothing$, since conflict sets are minimal $body(s)^- = \{\neg p\}$, and it must therefore be the case that $\neg p \in \mathcal{U}_{\Pi,\lambda-1}$. By the inductive hypothesis, $\neg p \in \mathcal{U}_{D,WF}$. Since $body(r) \subseteq \mathcal{T}_{D,WF}$ and $\neg p \in \mathcal{U}_{D,WF}$, then by Lemma 5.5, $p \in \mathcal{T}_{D,WF}$.

If $p \in \mathcal{T}_{\Pi,\lambda}$ and λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{\Pi,\kappa}$. By the inductive hypothesis, $p \in \mathcal{T}_{D,WF}$.

(Case 2) Let $p \in \mathcal{U}_{\Pi,\lambda}$ and suppose that λ is a successor ordinal. $\mathcal{U}_{\Pi,\lambda}$ is by definition unfounded relative to Π and $\mathcal{I}_{\Pi,\lambda-1}$. Let a be any literal such that $a \in \mathcal{U}_{\Pi,\lambda}$ and let $r' \in trans(r)$ for some $r \in R[a]$. Suppose r is strict. Since U_Π is monotone and $\mathcal{U}_{\Pi,\lambda}$ is unfounded relative to $\mathcal{I}_{\Pi,\lambda-1}$, it follows that there exists a classical $b \in body(r')$ such that $b \in \mathcal{U}_{\Pi,\lambda}$. Thus there exists a $b \in body(r)$ such that $b \in \mathcal{U}_{\Pi,\lambda}$.

Suppose r is defeasible. Then either (1) there exists a classical $b \in body(r')$ such that $b \in \mathcal{U}_{\Pi,\lambda}$ or else (2) the default literal $not\ \neg a$ appears in $body(r')$ and $\neg a \in \mathcal{T}_{\Pi,\lambda-1}$. If (1), then $body(r) \cap \mathcal{U}_{\Pi,\lambda} \neq \varnothing$. If (2) then by the inductive hypothesis $\neg a \in \mathcal{T}_{D,WF}$, and so there must be a rule $s \in R[\neg a]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and either (2.1) s is strict or else (2.2) for all rules $t \in R[a]$ (including $t = r$), $body(t) \cap \mathcal{U}_{D,WF} \neq \varnothing$.

Generalizing on r, for each rule $r \in R_s[a]$, $body(r) \cap (\mathcal{U}_{\Pi,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$. For each $r \in R_d[a]$, either $body(r) \cap (\mathcal{U}_{\Pi,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$ or else there exists an $s \in R_s[\neg a]$ and $body(s) \subseteq \mathcal{T}_{D,WF}$. Generalizing on a, it can be seen that $\mathcal{U}_{\Pi,\lambda}$ constitutes an unfounded set *wrt* D and $\mathcal{I}_{D,WF}$. Consequently, $\mathcal{U}_{\Pi,\lambda} \subseteq U_D(\mathcal{U}_{D,WF})$. Since $\mathcal{I}_{D,WF}$ is a fixpoint of W_D, we have $\mathcal{U}_{\Pi,\lambda} \subseteq \mathcal{U}_{D,WF}$, and hence $p \in \mathcal{U}_{D,WF}$.

If $p \in \mathcal{U}_{\Pi,\lambda}$ and λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{\Pi,\kappa}$. By the inductive hypothesis, $p \in \mathcal{U}_{D,WF}$. □

5.3 Translating Logic Programs into Defeasible Logic

Normal logic programs can also be translated into defeasible theories so that the assertions of the theory match the well-founded model. We show this by first translating the logic program into an equivalent one in which the default literals are relocated (*i.e.*, 'not p' is replaced with '$\neg p$' in existing rules, and new rules relating $not\ p$ and $\neg p$ are added). We then show the equivalence between defeasible logic and these programs.

5.3.1 Explicit Normal Programs

Definition 5.7 (Explicit Normal Programs)**.** *Let* Π *be a normal program. The explicit version of* Π *is the smallest extended program* Φ *such that*

(1) If $p\ :-\ a_1, \ldots, a_n,\ not\ b_1, \ldots,\ not\ b_m$ *appears in* Π*, then*

$p\ :-\ a_1, \ldots, a_n, \neg b_1, \ldots, \neg b_m$ *appears in* Φ*.*

(2) For each $p \in B_\Pi$*,* $\neg p\ :-\ not\ p$ *appears in* Φ*.*

Lemma 5.8. *Let* Π *be a normal program and* Φ *the explicit version of* Π*. For any* $b \in B_\Pi$ *and any ordinal* $\lambda \geq 0$*,*

(1) $\neg b \in \mathcal{T}_{\Phi,\lambda}$ *iff there exists a* $\kappa < \lambda$ *such that* $b \in \mathcal{U}_{\Phi,\kappa}$*, and*

(2) $\neg b \in \mathcal{U}_{\Phi,\lambda}$ *iff there exists a* $\kappa < \lambda$ *such that* $b \in \mathcal{T}_{\Phi,\kappa}$*.*

Proof.

(1) Let $\neg b \in \mathcal{T}_{\Phi,\lambda}$ for some ordinal λ. Then there exists a least successor ordinal $\kappa \leq \lambda$ such that $\neg b \in \mathcal{T}_{\Phi,\kappa}$. As $\neg b\ :-\ not\ b$ is the only rule with head $\neg b$, it must be the case that $b \in \mathcal{U}_{\Phi,\kappa-1}$.

Now suppose there is an ordinal $\kappa < \lambda$ such that $b \in \mathcal{U}_{\Phi,\kappa}$. Since $\neg b\ :-\ not\ b$ is a rule in Φ, it must be the case that $\neg b \in \mathcal{T}_{\Phi,\kappa+1}$. Either $\kappa + 1 = \lambda$, or else by monotonicity of $(\mathcal{I})$ we have $\neg b \in \mathcal{T}_{\Phi,\lambda}$.

(2) Suppose $\neg b \in \mathcal{U}_{\Phi,\lambda}$. Then there exists a least successor ordinal $\kappa \leq \lambda$ such that $\neg b \in \mathcal{U}_{\Phi,\kappa}$. Since $\neg b\ :-\ not\ b$ is the only rule in Φ with head $\neg b$, it must be the case that $b \in \mathcal{T}_{\Phi,\kappa-1}$.

Now suppose there is an ordinal $\kappa < \lambda$ such that $b \in \mathcal{T}_{\Phi,\kappa}$. As $\neg b \ :- \ not\ b$ is the only rule of Φ with head b, it must be the case that $b \in \mathcal{U}_{\Phi,\kappa+1}$. Either $\kappa + 1 = \lambda$, or else by monotonicity we have $b \in \mathcal{U}_{\Phi,\lambda}$. □

Lemma 5.9. *Let* Π *be a normal program and* Φ *the explicit version of* Π. *For all atoms* $p \in B_\Pi$, $\Pi \mid\!\approx_{WFS} p$ *iff* $\Phi \mid\!\approx_{WFS} p$, *and* $\Pi \not\approx_{WFS} p$ *iff* $\Phi \not\approx_{WFS} p$.

Proof. **(LR)** The proof is by induction on the sequence $\mathcal{I}_{\Pi,0}, \mathcal{I}_{\Pi,1}, \ldots$. Suppose for all $k < \lambda$ and all $p \in B_\Pi$, if $p \in \mathcal{T}_{\Pi,k}$, then $p \in \mathcal{T}_{\Phi,WF}$; if $p \in \mathcal{U}_{\Pi,k}$ then $p \in \mathcal{U}_{\Phi,WF}$.

(Case 1) Suppose $p \in B_\Pi$, $p \in \mathcal{T}_{\Pi,\lambda}$, and λ is a successor ordinal. Then there is a rule r with head p such that $body(r)^+ \subseteq \mathcal{T}_{\Pi,\lambda-1}$ and $body(r)^- \subseteq \mathcal{U}_{\Pi,\lambda-1}$. Let r' be the rule of Φ corresponding to r. By the inductive hypothesis, $body(r)^+ \subseteq \mathcal{T}_{\Phi,WF}$ and $body(r)^- \subseteq \mathcal{U}_{\Phi,WF}$. By Lemma 5.8 for each $q \in body(r)^-$ it follows that $\neg q \in \mathcal{T}_{\Phi,WF}$. Since $body(r)+ \subseteq \mathcal{T}_{\Phi,WF}$ and $q \in body(r)^-$ we have $\neg q \in \mathcal{T}_{\Phi,WF}$, it must be the case that $body(r') \subseteq \mathcal{T}_{\Phi,WF}$. Since r' is strict, then by definition of T_Φ and $\mathcal{I}_{\Phi,WF}$, $p \in \mathcal{T}_{\Phi,WF}$.

If λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{\Pi,\kappa}$. By the inductive hypothesis, $p \in \mathcal{T}_{\Pi,WF}$.

(Case 2) Suppose $p \in B_\Pi$, $p \in \mathcal{U}_{\Pi,\lambda}$, and λ is a successor ordinal. Let $q \in B_\Pi$ be any literal such that $q \in \mathcal{U}_{\Pi,\lambda}$. Then for all rules $r \in R_\Pi[q]$ there is a classical literal $a \in body(r)$ such that $a \in \mathcal{U}_{\Pi,\lambda}$ or else a default literal $not\ b$ such that $b \in \mathcal{T}_{\Pi,\lambda-1}$. If $b \in \mathcal{T}_{\Pi,\lambda-1}$, then by the inductive hypothesis $b \in \mathcal{T}_{\Phi,WF}$ and so from Lemma 5.8 $\neg b \in \mathcal{U}_{\Phi,WF}$. Generalizing, for all rules r, each corresponding rule r' has a classical literal $a \in body(r')$ such that $a \in \mathcal{U}_{\Pi,\lambda}$ or else a (still classical literal) $\neg b$ such that $\neg b \in \mathcal{U}_{\Phi,WF}$. Generalizing on q, it can be seen that $\mathcal{U}_{\Pi,\lambda}$ is an unfounded set *wrt* Φ and $\mathcal{I}_{\Phi,WF}$. As such, $\mathcal{U}_{\Pi,\lambda} \subseteq U_\Phi(\mathcal{I}_{\Phi,WF}) = \mathcal{U}_{\Phi,WF}$, and so $p \in \mathcal{U}_{\Phi,WF}$.

If λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{\Pi,\kappa}$. By the inductive hypothesis, $p \in \mathcal{U}_{\Phi,WF}$.

(RL) The proof is by induction on the sequence $\mathcal{I}_{\Phi,0}, \mathcal{I}_{\Phi,1}, \ldots$. Suppose for all $\kappa < \lambda$ and $p \in B_\Pi$, if $p \in \mathcal{T}_{\Phi,\kappa}$, then $p \in \mathcal{T}_{\Pi,WF}$; if $p \in \mathcal{U}_{\Phi,\kappa}$ then $p \in \mathcal{U}_{\Pi,WF}$.

(Case 1) Suppose $p \in B_\Pi$, $p \in \mathcal{T}_{\Phi,\lambda}$ and λ is a successor ordinal. Then there is a rule r' with head p such that $body(r) \subseteq \mathcal{T}_{\Phi,\lambda-1}$ (note that r' cannot contain default literals). By Lemma 5.8, for each $\neg b \in body(r')$, we have $b \in \mathcal{U}_{\Phi,\eta}$ for some $\eta < \lambda$. Let r be the rule of Π corresponding to r'. By the inductive hypothesis, $body(r)^+ \subseteq \mathcal{T}_{\Pi,WF}$ and $body(r)^- \subseteq \mathcal{U}_{\Pi,WF}$. By definition of T_Π and $\mathcal{I}_{\Pi,WF}$, $p \in \mathcal{T}_{\Pi,WF}$.

If λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{\Phi,\kappa}$. By the inductive hypothesis, $p \in \mathcal{T}_{\Pi,WF}$.

(Case 2) Now suppose $p \in B_\Pi$, $p \in \mathcal{U}_{\Phi,\lambda}$ and let q be any literal such that $q \in \mathcal{U}_{\phi,\lambda}$. If q is a classical negative literal, then no rules for q appear in Π. Suppose q is an atom. Since $q \in \mathcal{U}_{\Phi,\lambda}$, for all rules r' with head q there is a classical literal $a \in body(r)$ such that $a \in \mathcal{U}_{\Phi,\lambda}$. If a is of the form $\neg b$, then by Lemma 5.8, $b \in \mathcal{T}_{\Phi,\eta}$ for some $\eta < \lambda$. By the inductive hypothesis, each such b is in $\mathcal{T}_{\Pi,WF}$. Recall that if $\neg b$ appears in the body of r', then $not\ b$ appears in the corresponding rule r of Π. Generalizing on r', every rule r in Π with head p has a classical subgoal $a \in \mathcal{U}_{\Phi,\lambda}$ or else a default subgoal $not\ b$ such that $b \in \mathcal{T}_{\Pi,WF}$. Generalizing on q, it can be seen that $\mathcal{U}_{\Phi,\lambda}$ is an unfounded set *wrt* Π and $\mathcal{I}_{\Pi,WF}$. As such, $\mathcal{U}_{\Phi,\lambda} \subseteq U_\Pi(\mathcal{U}_{\Pi,WF}) = \mathcal{U}_{\Pi,WF}$, and so $p \in \mathcal{U}_{\Pi,WF}$.

If λ is a limit ordinal, then there exists a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{\Phi,\kappa}$. By the inductive hypothesis, $p \in \mathcal{U}_{\Pi,WF}$. □

5.3.2 Translation into Defeasible Theories

The relationship between Π and Φ makes the translation of Π into a defeasible theory D_Π apparent.

Definition 5.10 (Translation into Defeasible Theory)**.** *Let Π be a normal logic program. If rule r*

$$p\ :\!-\ a_1, \ldots, a_n,\ not\ b_1, \ldots,\ not\ b_m.$$

appears in Π, then r_{D_Π} is the rule

$$\{a_1, \ldots, a_n, \neg b_1, \ldots, \neg b_m\} \rightarrow p$$

Let $Str = \{r_{D_\Pi} | \, r \in \Pi\}$ *and* $Pr = \{\{\} \Rightarrow \neg p | \, p \in B_\Pi\}$. D_Π*, the defeasible theory translation of* Π*, is*

$$D_\Pi = \langle Str \cup Pr, C, \varnothing \rangle$$

where C *is the minimal conflict sets defined over the atoms of* Π.

The default literals in the program have become presumptions in the defeasible theory. The rules of the original program are strict in the defeasible theory. Translating D_Π back into a logic program using the Brewka inspired scheme yields Φ. Given the soundness (5.4) and completeness (5.6) theorems of the last section and also Lemma 5.8, it follows that p is provable in D_Π in ADL if and only if $\neg p$ is refutable in ADL, and $\neg p$ is provable in ADL if and only if p is refutable in ADL.

Theorem 5.11. *Let* Π *be a normal logic program and* D_Π *its defeasible theory translation. For any* $p \in B_\Pi$, $D_\Pi \mid\!\approx_{ADL} p$ *iff* $D_\Pi \not\approx_{ADL} \neg p$, *and* $D_\Pi \mid\!\approx_{ADL} \neg p$ *iff* $D_\Pi \not\approx_{ADL} p$.

Proof. Let Φ be the explicit form of Π. It is clear that translating D_Π into a logic program using Brewka's scheme yields Φ.

Suppose $D_\Pi \mid\!\approx_{ADL} p$. Then $\Phi \mid\!\approx_{WFS} p$ by Theorem 5.4 (Soundness). By Lemma 5.8, $\Phi \not\approx_{WFS} \neg p$. By Theorem 5.6 (Completeness) we have $D_\Pi \not\approx_{ADL} \neg p$. Suppose $D_\Pi \mid\!\approx_{ADL} \neg p$. Then $\Phi \mid\!\approx_{WFS} \neg p$ by 5.4. By Lemma 5.8, $\Phi \not\approx_{WFS} p$. By Theorem 5.6 we have $D_\Pi \not\approx_{ADL} p$.

Suppose $D_\Pi \not\approx_{ADL} p$. Then $\Phi \not\approx_{WFS} p$ by Theorem 5.4. By Lemma 5.8, $\Phi \mid\!\approx_{WFS} \neg p$. By Theorem 5.6 we have $D_\Pi \mid\!\approx_{ADL} \neg p$. Suppose $D_\Pi \not\approx_{ADL} \neg p$. Then $\Phi \not\approx_{WFS} \neg p$ by 5.4. By Lemma 5.8, $\Phi \mid\!\approx_{WFS} p$. By Theorem 5.6 we have $D_\Pi \mid\!\approx_{ADL} p$. □

Furthermore, Since Π and Φ yield the same consequences for the atoms of Π (Lemma 5.9) the assertions of D_Π agree with the well-founded model of Π *wrt* B_Π.

Theorem 5.12. *Let* Π *be a normal program and* D_Π *its defeasible logic translation. For any atom* $p \in B_\Pi$, $D_\Pi \mid\!\approx_{ADL} p$ *iff* $\Pi \mid\!\approx_{WFS} p$, *and* $D_\Pi \not\approx_{ADL} p$ *iff* $\Pi \not\approx_{WFS} p$.

Proof. Suppose $D_\Pi \mid\!\approx_{ADL} p$. Then $\Phi \mid\!\approx_{WFS} p$ by 5.4. By Lemma 5.9, $\Pi \mid\!\approx_{WFS} p$. Suppose $D_\Pi \not\mid\!\approx_{ADL} p$. Then $\Phi \not\mid\!\approx_{WFS} p$ by 5.4. By Lemma 5.9, $\Pi \not\mid\!\approx_{WFS} p$.

Suppose $\Pi \mid\!\approx_{WFS} p$. Then $\Phi \mid\!\approx_{WFS} p$ by Lemma 5.9. By Theorem 5.6 we have $D_\Pi \mid\!\approx_{ADL} p$. Suppose $\Pi \not\mid\!\approx_{WFS} p$. Then $\Phi \not\mid\!\approx_{WFS} p$ by Lemma 5.9. By Theorem 5.6 we have $D_\Pi \not\mid\!\approx_{ADL} p$. □

Example 5.13. *A logic program* Π*, its explicit form* Φ*, and its defeasible logic translation* D_Π *are shown below.*

$$
\begin{array}{lllll}
\Pi= & \{p \ :- \ not\ q, & q \ :- \ not\ p\} & & \\
\Phi= & \{p \ :- \neg q, & q \ :- \neg p, & \neg p \ :- \ not\ p, & \neg q \ :- \ not\ q\} \\
D_\Pi = & \langle\{\{\neg q\} \rightarrow p, & \{\neg p\} \rightarrow q, & \{\} \Rightarrow \neg p, \{\} \Rightarrow \neg q\}, & C,\varnothing\rangle
\end{array}
$$

In the rules of Π, we have replaced each subgoal $not\ p$ (alternatively $not\ q$) with $\neg p$ (alternatively $\neg q$) and added the rules $\neg p \ :- \ not\ p$ and $\neg q \ :- \ not\ q$. The explicitly negative literals $\neg p$ and $\neg q$ occur nowhere in the original program Π and so it is safe to equate, *e.g.*, $not\ p$ with $\neg p$. The well-founded model of both Π and Φ is empty. In D_Π, the presumption of $\neg p$ prevents concluding $\neg q$, but there is no superior argument for q, and so $\neg q$ is not refuted, either. The same holds for $\neg p$, and because of this nothing can be determined for p or q. The set of assertions of D_Π is empty.

5.3.3 Extended Logic Programs into Defeasible Theories

The above translation of logic program to defeasible theory specified that the logic program is *normal*. In an extended logic program, a classically negative literal $\neg p$ might already appear in Π, and so the maneuver of replacing $not\ p$ with $\neg p$ is no longer acceptable. The equivalence of Φ and Π would in general no longer hold.

Example 5.14.

Π	Φ
$p.$	$p.$
$\neg p.$	$\neg p.$
$q \ :- \ not \ p.$	$q \ :- \ \neg p.$
	$\neg p \ :- \ not \ p.$
	$\neg q \ :- \ not \ q.$

Using the original well-founded semantics, the well-founded model of Π is $\langle\{p, \neg p\}, \{q, \neg q\}\rangle$. However, the well-founded model of Φ is $\langle\{p, \neg p, q\}, \{\neg q\}\rangle$. If we instead replace $not \ p$ with some entirely new literal neg_p, then the equivalence is—*in a sense*—restored. However, in the defeasible logic translation, p and $\neg p$ do not conflict (the program rule $neg\_p \ :- \ not \ p$ would be interpreted as a $\{\} \Rightarrow neg\_p$ and neg_p and p would conflict). This at first seems very odd. However, strictly speaking, in the WFS there is really no connection between an atom p and its classical negation $\neg p$, either. The literal $\neg p$ is just a "rather strange looking atom".

Example 5.15. *Let* Π *be the below program.*

(1) $\neg p \ :- \ not \ q.$
(2) $\neg p \ :- \ not \ p.$

Π *can be translated to program* Φ*:*

(1) $\neg p \ :- \ neg\_q.$
(2) $\neg p \ :- \ neg\_p.$
(3) $neg\_q \ :- \ not \ q.$
(4) $neg\_p \ :- \ not \ p.$

which in turn becomes the below defeasible theory D_Π*, in which* $\{neg\_p, p\}$ *and* $\{neg\_q, q\}$ *are both conflict sets (and* $\{\neg p, p\}$ *and* $\{\neg q, q\}$ *are* not*).*

(1) $\{neg\_q\} \rightarrow \neg p$
(2) $\{neg\_p\} \rightarrow \neg p$
(3) $\{\} \Rightarrow neg\_q$
(4) $\{\} \Rightarrow neg\_p$

The well-founded model of Π is $\langle\{\neg p\}, \{p, q\}\rangle$. The model according to D_Π is $\langle\{\neg p, neg\_p, neg\_q\}, \{p, q\}\rangle$ (since no rules for p or q appear in the defeasible theory, they are obviously unfounded). As in the case for normal logic programs, translating D_Π into a logic program yields Φ, and the consequences of D_Π under ADL correspond to the well-founded model of Φ.

5.4 Alternating Fixpoints: Correspondence with the WFS

Under the translation of defeasible theories into logic programs, α corresponds to the GL-operator γ used for the stable model/answer set and WFS semantics. The restriction that $C[p]$ be finite for each $p \in Lit_D$ is still needed, however, to prevent rules in the logic program from being infinite in length. Since it has already been shown that the semantics of ADL corresponds to the WFS for a restricted class of defeasible theories, it follows that α also can be used to characterize the consequences under ADL of defeasible theories in this class.

Lemma 5.16. *If $D = \langle R_{sd}, C, \varnothing\rangle$ is a defeasible theory such that $C[p]$ is finite for each $p \in Lit_D$, and Π is the logic program translation of D, then for any $X \subseteq Lit_D$, $\alpha_D(X) = \gamma_\Pi(X)$.*

Proof. The proof proceeds by induction on the simple immediate consequence operator T used to compute the closure of reducts. Note that this operator is continuous and $T^\omega = \bigcup_{n<\omega} T^n$, and so it suffices to show that for each $n < \omega$, $T^n_{\Pi^X}(\bot) = T^n_{D^X}(\bot)$. Note that $T^0_{D^X}(\bot) = T^0_{\Pi^X}(\bot) = \bot$ and so the claim holds for $i = 0$. Suppose it holds for $i < n$, and let $p \in T^n_{D^X}(\bot)$. Then there is an $r \in R_{s,d}[p]$ in the reduct D^X such that $body(r) \subset T^{n-1}_{D^X}(\bot)$. If r is strict, then there exists an $r' \in \Pi_D$ such that $r' = r$. Since r' is strict, $r' \in \Pi^X_D$. Since $body(r) \subset T^{n-1}_{D^X}(\bot)$, by the inductive hypothesis $body(r) \subset T^{n-1}_{\Pi^X}(\bot)$, and so by definition of the immediate consequence operator $p \in T^n_{\Pi^X}(\bot)$. If r is defeasible, then for all $c \in C[p]$, there exists a $q \in (c - \{p\})$ such that $q \notin X$. As such, there exists a rule $r' \in trans(r)$ such that for each $not\ b$ in the body of r' we have $b \notin X$. As this is so, $r' \in \Pi^X_D$ and every default literal in the body of r' has been deleted. Since $body(r) \subseteq T^{n-1}_{D^X}(\bot)$, by inductive hypothesis $body(r) \subseteq T^{n-1}_{\Pi^X}(\bot)$. Since $body(r) = body(r')^+$ and all default literals in $body(r')$ have been deleted, it follows that $p \in T^n_{\Pi^X}(\bot)$.

Now suppose $p \in T^n_{\Pi^X}(\bot)$. Then there is a rule t in Π^X such that $body(t) \subset T^{n-1}_{\Pi^X}(\bot)$. If t corresponds to a strict rule of D, then $t \in D^X$ and by inductive hypothesis $body(t) \subset T^{n-1}_{D^X}(\bot)$, and so by definition of the immediate consequence operator $p \in T^n_{D^X}(\bot)$. If t corresponds to a defeasible rule t', since t appears in the reduct of Π *wrt* X it must be the case that every default literal of t has been deleted. This means that for each conflict set $c \in C[p]$ there is an element $q \in c - \{p\}$ such that $q \notin X$. Since this is so, then by definition of D^X, $t' \in D^X$. By the inductive hypothesis, $body(t') \subset T^{n-1}_{D^X}(\bot)$ and so as before, $p \in T^n_{D^X}(\bot)$. □

Previously, we showed a correspondence between the well-founded model for logic programs and the well-founded model for ADL; this correspondence holds for theories with minimal conflict sets and no defeaters or precedence. Given the correspondence just shown between α and the GL-operator γ, we can now see that α can be used to determine the consequences of ADL for these theories.

Furthermore, since the Gelfond-Lifschitz operator γ is used to define the stable models of a program Π, we can use α to define a similar semantics for defeasible theories.

Definition 5.17. *Let $D = \langle R, C, \prec \rangle$ be a defeasible theory such that $R_u = \varnothing$ and $S \subseteq Lit_D$. Then S is a* stable set *of D iff $S = \alpha_D(S)$.*

Example 5.18.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$
(3) $\{p\} \Rightarrow q$
(4) $\{\} \Rightarrow \neg q$

The stable sets of this example are $\{p, q\}$, $\{p, \neg q\}$, and $\{\neg p, \neg q\}$.

5.5 Eliminating Precedence and Defeaters

The correspondences which exists between ADL and the well-founded semantics for logic programs assume (1) that the precedence relation of the defeasible theory is empty (2) no defeaters

are used, and (3) conflict sets are not extended. It turns out that these features can be eliminated from defeasible theories without limiting its expressiveness. We present a method of translating a general defeasible theory D into an equivalent one in which $R_u = \varnothing$, $\prec = \varnothing$, and C is minimal. The translation works for both NDL and ADL, but again we must insist that for all $p \in Lit_D$, $C[p]$ is finite. Even under this restriction, it must be stressed that the computational space penalty for removing these elements is rather severe—the resulting theory is exponentially larger than the original.

The general idea behind the translation is to introduce new literals of the form $supported(r)$ and $fires(r)$ which explicitly encode when the head of a rule r of D is supported and when it may be detached (or "fire"). If r is strict, then its head may be detached if and only if its body is supported. If r is defeasible, however, we want it to be possible for r to be supported but still not fire. Specifically, a competing rule with head $\neg fires(r)$ prevents it from firing. The conflict sets of D are used to create these competing rules.

Definition 5.19 (Normal Form)**.** *Let D be a defeasible theory such that for all $p \in Lit_D$, $C[p]$ is finite. The normal form of D is the smallest theory E such that:*

(1) For each strict r: $A \rightarrow p$ in D, the following rules appear in E:

(1.1) r': $A \rightarrow supported(r)$.
(1.2) r'': $\{supported(r)\} \rightarrow fires(r)$.
(1.3) r''': $\{fires(r)\} \rightarrow p$.

(2) For each defeasible rule r: $A \Rightarrow p$ in D, the following rules appear in E:

(2.1) r': $A \rightarrow supported(r)$.
(2.2) r'': $\{supported(r)\} \Rightarrow fires(r)$.
(2.3) r''': $\{fires(r)\} \rightarrow p$.

(3) For each defeater rule r: $A \rightsquigarrow p$ in D, the following rules occur in E:

(3.1) r': $A \rightarrow supported(r)$.
(3.2) r'': $\{supported(r)\} \Rightarrow fires(r)$.

(4) Let $c = \{q_1, q_2, \ldots, q_n, p\} \in C[p]$ be a conflict set, $r \in R_d[p]$ and $s_1, s_2, \ldots, s_n$ rules such that $s_i \in R[q_i]$ and $s_i \not\prec r$.

(4.1) If for all s_i, $s_i \in R_s$ or $r \prec s_i$, the following rule appears in E:

$$\{supported(s_1), supported(s_2), \ldots, supported(s_n)\} \rightarrow \neg fires(r)$$

(4.2) Otherwise, the following rule appears in E:

$$\{supported(s_1), supported(s_2), \ldots, supported(s_n)\} \Rightarrow \neg fires(r)$$

In item 4, a rule s_i is considered only if it is not inferior to r, for the reason that an inferior rule cannot be used to defeat r. For any conflict set $c \in C[head(r)]$, we will use $trans(c, r)$ to denote the set of rules for $\neg fires(r)$ created from conflict set c. Importantly, the conflict sets of the resulting theory are not closed under strict rules. For any literal p, $\{p, \neg p\}$ is the only conflict set containing p. Thus the above scheme can be used as an alternative means of embedding DL into logic programming under the WFS. Note that this scheme allows the incorporation of all features of defeasible logic (save an infinite $C[p]$).

If D itself uses minimal conflict sets, then if $r \in R_{sd}[p]$ and $s \in R[\neg p]$, condition 4 above reduces to $\{supported(s)\} \rightarrow \neg fires(r)$ if s is strict or $r \prec s$, and to $\{supported(s)\} \Rightarrow \neg fires(r)$ if $s \not\prec r$. Note that under this restriction the translation *is* modular and is polynomial in the size of the original theory D. Let L be the number of distinct literals appearing as rule-heads in the program. Suppose on average there are R rules for each literal p, and each rule consists of A occurrences of literals. Then there are in total $(L \times R)$ rules in the program, and the size of the full program is $(L \times R) \times A$. Given the translation, each rule has at most R rules that conflict with it, and each of these is two literals long: $(L \times R) \times R \times 2$. So the total size of the translated program is: $((L \times R) \times A) + ((L \times R) \times R \times 2) = (L \times R)(A + 2R)$.

5.5.1 Proof of Equivalence of D and E

Theorem 5.20. *If D is a defeasible theory such that for all $p \in Lit_D$, $C[p]$ is finite, and E is the normal form of D, then $p \in \mathcal{T}_{D,WF}$ iff $p \in \mathcal{T}_{E,WF}$, and $p \in \mathcal{U}_{D,WF}$ iff $p \in \mathcal{U}_{E,WF}$.*

Proof. The proof proceeds by induction on the sequence of interpretations $\mathcal{I}_{D,0}, \mathcal{I}_{D,1}, \ldots$ (and also $\mathcal{I}_{E,0}, \mathcal{I}_{E,1}, \ldots$).

(L to R): We will prove that for all $\kappa \geq 0$, if $p \in \mathcal{T}_{D,\kappa}$ then $p \in \mathcal{T}_{E,WF}$, and if $p \in \mathcal{U}_{D,\kappa}$ then $p \in \mathcal{U}_{E,WF}$. The claim holds trivially for $\kappa = 0$. Suppose that it holds for all $\kappa < \lambda$.

(Case 1) Suppose $p \in \mathcal{T}_{D,\lambda}$ and λ is a successor ordinal. Then there exists an $r \in R[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$ and either (1) $r \in R_s$ or else (2) $r \in R_d$ and for each $c \in C[p]$, there exists a $q \in c - \{p\}$ and for all $s \in R[q]$, $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$ or $s \prec r$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{E,WF}$.

Suppose (1) holds. E then contains the below strict rules; by examining these, it is clear that $p \in \mathcal{T}_{E,WF}$.

(1) r': $body(r) \rightarrow supported(r)$.
(2) r'': $supported(r) \rightarrow fires(r)$.
(3) r''': $fires(r) \rightarrow p$.

Suppose (2) obtains. E contains the below rules.

(1) r': $body(r) \rightarrow supported(r)$.
(2) r'': $supported(r) \Rightarrow fires(r)$.
(3) r''': $fires(r) \rightarrow p$.

Since $body(r) \subseteq \mathcal{T}_{E,WF}$, $supported(r) \in \mathcal{T}_{E,WF}$. Also, by the inductive hypothesis, for every conflict set $c \in C[p]$, there is an element $q \in c - \{p\}$ such that for every rule $s \in R[q]$, either (i) $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$, or else (ii) $s \prec r$. In other words, if $s \not\prec r$ then $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$. Note that if $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$, then $supported(s) \in \mathcal{U}_{E,WF}$. For each $t \in trans(c, r)$ and each $supported(a) \in body(t)$, by definition of $trans(c, r)$ it must be the case that $a \not\prec r$. Since each rule $t \in trans(c, r)$ contains a $supported(s)$ such that $s \in R[q]$, and since $s \not\prec r$, it follows that for every rule $t \in trans(c, r)$, $body(t) \cap \mathcal{U}_{E,WF} \neq \varnothing$. Generalizing on c, for every rule $t \in R_E[\neg fires(r)]$, $body(t) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Thus we have $supported(r) \subseteq \mathcal{T}_{E,WF}$ and for each $t \in R_E[\neg fires(r)]$, $body(t) \cap \mathcal{U}_{E,WF} \neq \varnothing$. By definition of T_E and $\mathcal{I}_{E,WF}$, both $fires(r) \in \mathcal{T}_{E,WF}$ and $p \in \mathcal{T}_{E,WF}$.

If λ is a limit ordinal, then if $p \in \mathcal{T}_{D,\lambda}$ there must be a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{D,\kappa}$. By the inductive hypothesis we have $p \in \mathcal{T}_{E,WF}$.

(Case 2) Now suppose that $p \in \mathcal{U}_{D,\lambda}$ and λ is again a successor ordinal. By definition of unfounded sets for ADL (alternatively, NDL), we have

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U}_{D,\lambda-1} \cup \mathcal{U}_{D,\lambda}) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

  (2.1) $body(r) \cap (\mathcal{U}_{D,\lambda-1} \cup \mathcal{U}_{D,\lambda}) \neq \varnothing$, or
  (2.2) there is a $c \in C[p]$ such that for each $q \in c - \{p\}$ there is a rule $s \in R[q]$ such that

    (2.2.1) $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ and
    (2.2.2) $r \prec s$ or s is strict. (for NDL, $s \not\prec r$)

Since (by monotonicity of U_D) $\mathcal{U}_{D,\lambda-1} \subseteq \mathcal{U}_{D,\lambda}$, conditions 1 and 2.1 can be simplified to '$body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$'.

Let $X = \mathcal{U}_{D,\lambda} \cup \{fires(r) |\ r \in R_D[p]$ and $p \in \mathcal{U}_{D,\lambda}\} \cup \{supported(r), fires(r) |\ body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing\}$. We will show that X is unfounded *wrt* E and some interpretation $\mathcal{I}_{E,\alpha}$ in the sequence. Note that there are three types of literal that appear in X: Those that appear in $\mathcal{U}_{D,\lambda}$, those that have the form $fires(r)$, and those that have the form $supported(r)$. We examine each type in turn.

(Case 2.1) If $p \in X \cap \mathcal{U}_{D,\lambda}$, then by definition of X for all rules $fires(r) \rightarrow p \in R_E[p]$, $fires(r) \in X$. This exhausts all rules in E for p.

(Case 2.2) If $supported(r) \in X$, then by definition of X, $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$.

(Case 2.3) Suppose $fires(r) \in X$. Recall that $r \in R_D[p]$ for some p, and E contains only one rule with head $fires(r)$—namely, $r'' : supported(r) \dashrightarrow fires(r)$. Obviously, $body(r)$ intersects $\mathcal{U}_{D,\lambda}$ or it doesn't. If $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, then $supported(r) \in X$. If $body(r) \cap \mathcal{U}_{D,\lambda} = \varnothing$, then by definition of X it must be the case that $r \in R_D[p]$ for some $p \in \mathcal{U}_{D,\lambda}$. Since $p \in \mathcal{U}_{D,\lambda}$ and $body(r) \cap \mathcal{U}_{D,\lambda} = \varnothing$, r must be defeasible and so $supported(r) \Rightarrow fires(r)$ is the only rule for $fires(r)$ in E. Furthermore, there must exist a conflict set $c \in C[p]$ such that for all $q \in c - \{p\}$, there is a rule $s \in R[q]$ such that $body(s) \in \mathcal{T}_{D,\lambda-1}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$). By the inductive hypothesis, $body(s) \in \mathcal{T}_{E,WF}$. As such $supported(s) \in \mathcal{T}_{E,WF}$. Furthermore, since for each s, $r \prec s$ or s is strict (in NDL, $s \not\prec r$) the rule $t : supported(s_1), supported(s_2), \ldots, supported(s_\lambda) \rightarrow \neg fires(r)$ appears in E (for NDL this rule will be defeasible, not strict). And so there exists a rule $t \in R_E[\neg fires(r)]$ such that $body(t) \subseteq \mathcal{T}_{E,WF}$ and t is strict (for NDL, $t \not\prec r''$).

Generalizing on the literals of X, for all literals a in X, the following obtains: (1) for every strict rule $r \in R_s[a]$, $body(r) \cap (X \cup \mathcal{U}_{E,WF}) \neq \varnothing$; (2) for every defeasible rule, either $body(r) \cap (X \cup \mathcal{U}_{E,WF}) \neq \varnothing$ or else there exists a conflict set $c \in C[a]$ such that for each $b \in c - \{a\}$, there is a rule $s \in R[b]$ such that $body(s) \subseteq \mathcal{T}_{E,WF}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$). Thus X is an unfounded set *wrt* E and $\mathcal{I}_{E,WF}$. As such $X \subseteq U_E(\mathcal{I}_{E,WF}) = \mathcal{U}_{E,WF}$, and so $p \in \mathcal{U}_{E,WF}$.

If λ is a limit ordinal, $p \in \mathcal{U}_{D,\lambda}$ implies the existence of a least successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{D,\kappa}$. By the inductive hypothesis, $p \in \mathcal{U}_{E,WF}$.

(R to L) We will prove that for all $\kappa \geq 0$, if $p \in \mathcal{T}_{E,\kappa}$ then $p \in \mathcal{T}_{D,WF}$, and if $p \in \mathcal{U}_{E,\kappa}$ then $p \in \mathcal{U}_{D,WF}$. The claim holds trivially for $\kappa = 0$. Suppose that it holds for all $\kappa < \lambda$.

(Case 1) Suppose $p \in \mathcal{T}_{E,\lambda}$. Then there exist rules $r''' : fires(r) \rightarrow p \in R_{sd}[p]$ and $r'' : supported(r) \dashrightarrow fires(r)$ and r': $body(r) \rightarrow supported(r)$ and a least ordinal $\eta < \lambda$ such that $fires(r) \in \mathcal{T}_{E,\eta}$, $supported(r) \in \mathcal{T}_{E,\eta}$, and $body(r) \subseteq \mathcal{T}_{E,\eta}$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D,WF}$. Rule r is either strict or defeasible. We treat each case in turn. If $r \in R_s[p]$, then each of the above rules is strict and so clearly $p \in \mathcal{T}_{D,WF}$.

Suppose $r \in R_d[p]$. Then for every rule $t \in R[\neg fires(r)]$, there exists a $supported(s) \in body(t)$ such that $supported(s) \in \mathcal{U}_{E,\eta}$. This implies that $body(s) \cap \mathcal{U}_{E,\eta} \neq \varnothing$. By the inductive hypothesis, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Let c be some conflict set of D containing p. Then for each $t \in trans(c, r)$ there is a $supported(s) \in body(t)$ such that $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Given the manner in which the rules of $trans(c, r)$ are constructed, then there must be a $q \in c - \{p\}$ such that for each rule $s \in R[q]$, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or else $s \prec r$. Generalizing on c, for each conflict set $c \in C[p]$ there is a $q \in c - \{p\}$ such that for each rule $s \in R[q]$, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $s \prec r$. By definition of T_D and $\mathcal{I}_{D,WF}$, $p \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose $p \in \mathcal{U}_{E,\lambda}$ and let a be any literal in $(B_D \cap \mathcal{U}_{E,\lambda})$. Then for each rule r''': $fires(r) \rightarrow a$, $fires(r) \in \mathcal{U}_{E,\lambda}$ and for each rule r'': $supported(r) \dashrightarrow fires(r)$, either (1) r'' is strict or defeasible and $supported(r) \in \mathcal{U}_{E,\lambda}$, or (2) r'' is defeasible and there exists a rule $t : supported(s_1), supported(s_2), \ldots, supported(s_\lambda) \dashrightarrow \neg fires(r)$ such that $body(t) \subseteq \mathcal{T}_{E,\lambda-1}$ and t is strict (for NDL, $t \not\prec r''$). If (1) it follows that $body(r) \subseteq \mathcal{U}_{E,\lambda}$. If (2) then for

each $supported(s) \in body(t)$, $body(s) \subseteq \mathcal{U}_{E,\gamma}$ for some $\gamma < \lambda$ and by the inductive hypothesis $body(s) \subseteq \mathcal{U}_{D,WF}$. Given the manner in which rules such as t are created, there exists a conflict set $c \in C[a]$ such that for all $q \in c - \{a\}$ there is a rule $s \in R_D[q]$ such that $body(s) \in \mathcal{T}_{D,WF}$ and $r \prec s$ or s strict (for NDL, $s \not\prec r$).

Generalizing on a, for all $a \in (Lit_D \cap \mathcal{U}_{E,\lambda})$, then for all $r \in R_{sd}[a]$, either $body(r) \cap \mathcal{U}_{E,\lambda}$ or else $r \in R_d$ and there exists a conflict set $c \in C[a]$ and for all $q \in c - \{a\}$, there is a $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$). As such $\mathcal{U}_{E,\lambda}$ constitutes an unfounded set *wrt* D and $\mathcal{T}_{D,WF}$ and so $p \in \mathcal{U}_{D,WF}$. □

5.5.2 An Alternative Means of Eliminating Defeaters

This section presents an alternative and simpler method of eliminating defeaters from a defeasible theory.

Definition 5.21. *Let $D = \langle R_s \cup R_d \cup R_u, C_D, \prec \rangle$ be a defeasible theory. Define the defeater–free form of D to be $E = \langle R_s \cup R_d \cup S, C_E, \prec \rangle$, where*

(1) $S = \{r : A \Rightarrow p^* |\ r : A \rightsquigarrow p \in R_u[p]\}$.
(2) *C_E is defined inductively:*

- *(2.1) If $c \in C_D$, then $c \in C_E$.*
- *(2.2) If $c \in C_E[p]$ and $p \in Lit_D$, then $(c - \{p\}) \cup \{p^*\} \in C_E$.*
- *(2.3) Only those sets satisfying 2.1 or 2.2 are included in C_E.*

We have removed every defeater rule from D, replacing it with a new defeasible rule having as a head a new literal p^* (which stands in place of p). The new rules have the same labels and precedence as the old (and so the precedence relation has been left unaltered). From the definition for conflict sets in E, it can be seen that $C_D \subseteq C_E$, and from each $c \in C_D$ new conflict sets c' have been created by replacing one or more literals p with their starred counterparts p^*. Call c the *basis* for each c'. Note that the new literals p^* do not appear in the body of any rule of E, and so no such literal can be used as a witness of the success or failure of any rule. Furthermore, while C_E contains $2^{|C_D|}$ conflict sets, many of them can be pruned away, simply because some marked literals will have no supporting rules in R_E.

Example 5.22.

D	C_D	E	C_E
$\{\} \rightsquigarrow p$ $\{\} \Rightarrow \neg p$	$\{p, \neg p\}$	$\{\} \Rightarrow p^*$ $\{\} \Rightarrow \neg p$	$\{p, \neg p\}$, $\{p^*, \neg p\}$, $\{p, \neg p^*\}$, $\{p^*, \neg p^*\}$

In the above example, the undercutting defeater rule for p has been replaced with a defeasible rule for p^* and the conflict sets have been amended accordingly. The well-founded model for D according to ADL is $\langle\{\}, \{p\}\rangle$. For NDL, the model is $\langle\{\}, \{p, \neg p\}\rangle$. For E, the models are $\langle\{\}, \{p, p^*, \neg p^*\}\rangle$ and $\langle\{\}, \{p, p^*, \neg p, \neg p^*\}\rangle$, respectively. On the original literals of D, the models of D and E agree.

The proof that D and E correspond is shown below. The following technical lemma, which relates the rules of D and E, will be useful.

Lemma 5.23. *Let D be a defeasible theory and E its defeater–free form. If there exists a set $X \subseteq R_E$ such that*

(1) for each conflict set $x \in C_E[p]$ there exists a literal $a \in x - \{p\}$ such that $R_E[a] \subseteq X$, and
(2) $R_{D,sd} \cap X = R_{E,sd} \cap X$, and
(3) for all $r \in R_{D,u}$, $r \in X$ iff $r' \in X$.

then for each conflict set $y \in C_D[p]$ there exists a literal $b \in y - \{p\}$ such $R_D[b] \subseteq X$.

Proof. We will consider the rules of X 'marked'. Suppose that for each conflict set $x \in C_E[p]$ there exists a literal $a \in x - \{p\}$ such that $R_E[a] \subseteq X$. Suppose for a proof by contradiction that there exists a $c \in C_D[p]$ such that for all $q \in c - \{p\}$ $R_D[q] \not\subseteq X$. That is, for each $q_i \in c - \{p\}$ there is a rule $s_i \notin X$. We may separate the literals of $c - \{p\}$ into two classes: SD, the set of literals which have a strict or defeasible rule not in X; and U, the literals that have a defeater rule (but not a strict or defeasible one) not in X. We will construct the following set S of literals (this will be a conflict set of E): (1) add p to S; (2) for each $q \in SD$, add q to S; for all $q \in U$, add q^* to S. Note that if a defeater rule $r \in (R_D \cap X)$, then $r' \in (R_D \cap X)$, where r' is the corresponding defeasible rule in E. One can see that S is a conflict set of E. Furthermore, every literal of S has

a rule in E not in X. This contradicts our assumption, and so we infer that for every conflict set $y \in C_D[p]$, there exists a literal $b \in y - \{p\}$ such $R_D[b] \subseteq X$. □

Theorem 5.24. *Let D be a defeasible theory and E its defeater–free form. For every literal $p \in Lit_D$ and all $\lambda \geq 0$, $p \in \mathcal{T}_{D,\lambda}$ iff $p \in \mathcal{T}_{E,\lambda}$ and $p \in \mathcal{U}_{D,\lambda}$ iff $p \in \mathcal{U}_{E,\lambda}$.*

Proof. The claim above is satisfied for where $\lambda = 0$. Suppose $\lambda > 0$ and the claim holds for all $\kappa < \lambda$.

(L to R):

Suppose $p \in \mathcal{T}_{D,\lambda}$ and λ is a successor ordinal. Then there exists an $r \in R_D[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$ and either (1) $r \in R_{D,s}$ or (2) $r \in R_{D,d}$ and for all conflict sets $c \in C_D[p]$, there exists a $q \in c - \{p\}$ such that for all $s \in R_D[q]$, $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$ or $s \prec r$. By inductive hypothesis $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. For any such rule s, by the inductive hypothesis, if $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$ then $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. By definition of E, $r \in R_E$. If (1) holds, then by definition of T_E, $p \in \mathcal{T}_{E,\lambda}$. Suppose (2) holds and let $c \in C_D[p]$ be a conflict set and $q \in c - \{p\}$ a literal satisfying the above. Since $C_D \subseteq C_E$, $c \in C_E$. Since $R_E[q] \subseteq R_D[q]$, for all rules $s \in R_E[q]$, $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$. Let c' be any conflict set constructed from basis c such that q is replaced with q^*. By definition of E and q^*, for each defeater $s \in R_{D,u}[q]$ there is a corresponding $s' \in R_{E,d}[q^*]$ and $R_E[q^*] - R_{E,d}[q^*] = \varnothing$. Since for each $s \in R_D[q]$, $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$, it follows that for each $s' \in R_E[q^*]$, $body(E') \cap \mathcal{U}_{D,E-1} \neq \varnothing$ or $s' \prec r$. Generalizing on c and c', all conflict sets in $C_E[p]$ are accounted for, and so $p \in \mathcal{T}_{E,\lambda}$.

Now suppose $p \in \mathcal{U}_{D,\lambda}$. Then for all rules $r \in R_{D,s}[p]$, $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, and for all rules $r \in R_{D,d}[p]$, either $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$ or there exists a conflict set $c \in C_D[p]$ and for all $q \in c - \{p\}$, there exists a rule $s \in R_D[q]$ such that $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$). Let r be a defeasible rule for p and $c \in C_D[p]$ a conflict set as described above. We can construct a conflict set c' from c as follows: Include p in c'. If the rule s described above is in $R_{sd}[q]$, include q in c'. If $s \in R_{D,u}[q]$, include q^* in c'. Thus for each literal in $a \in c' - \{p\}$, there exists a $s \in R_E[a]$ such that $body(a) \subseteq \mathcal{T}_{D,\lambda-1}$ and $r \prec s$ or s is strict (for NDL, $r \not\prec s$). By

the inductive hypothesis for each s, $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$. Generalizing on p, it is clear that $\mathcal{U}_{D,\lambda}$ is unfounded *wrt* to E and $\mathcal{I}_{E,\lambda-1}$.

If λ is a limit ordinal, then if $p \in \mathcal{T}_{D,\lambda}$, there exists a least ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{D,\kappa}$. By the inductive hypothesis, we have $p \in \mathcal{T}_{E,\kappa}$. By monotonicity, $p \in \mathcal{T}_{E,\lambda}$. Similarly, if $p \in \mathcal{U}_{D,\lambda}$, there exists a least ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{D,\kappa}$. By the inductive hypothesis, we have $p \in \mathcal{U}_{E,\kappa}$. By monotonicity, $p \in \mathcal{U}_{E,\lambda}$.

(R to L):

Recall that for all literals $a \in Lit_D$, $R_{D,sd}[a] = R_{E,sd}[a]$. Let $p \in Lit_D$ and suppose $p \in \mathcal{T}_{E,\lambda}$ and λ is a successor ordinal. Then there exists an $r \in R_E[p]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$ and either (1) $r \in R_{E,s}$ or (2) $r \in R_{D,d}$ and for all conflict sets $c' \in C_E[p]$, there exists a literal $x \in c' - \{p\}$ such that for all $s \in R_E[x]$, $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$. Let X consist of rules of $R_D \cup R_E$ such that $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$. If (1), then by inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$ and $p \in \mathcal{T}_{D,\lambda}$ by definition of T_D. Suppose (2) holds. Then again by the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$. Since for all conflict sets $c' \in C_E[p]$ there exists a $x \in c' - \{p\}$ such that for all $s \in R_E[x]$, $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$, then by Lemma 5.23, for all conflict sets $c \in C_D[p]$ there exists a $q \in c - \{p\}$ such that for all $s \in R_D[q]$, $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $s \prec r$. By the inductive hypothesis, for any such s, if $s \not\prec r$ then $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. Since $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$ and each conflict set in $C[p]$ is accounted for, $p \in \mathcal{T}_{D,\lambda}$.

Now suppose $p \in \mathcal{U}_{E,\lambda}$ and λ is a successor ordinal. Then for all rules $r \in R_{E,s}[p]$, $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, and for all rules $r \in R_{E,d}[p]$, either $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$ or there exists a conflict set $c' \in C[p]$ and for all $q_i \in c' - \{p\}$, there exists a rule $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$). Let c be the basis of c'. Recall that for each literal $a \in c' - \{p\}$, there is a corresponding literal $q \in c - \{p\}$, and importantly, each rule $s \in R_E[a]$ has a corresponding rule in $R_D[q]$ with exactly the same priorities. And so for each $q \in c - \{p\}$ there is a $s \in R_D[q]$ such that $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$ and $r \prec s$ or s is strict (for NDL, $s \not\prec r$). By the inductive hypothesis, if $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$ then $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$. Generalizing on p, it can be seen that $\mathcal{U}_{E,\lambda}$ is an unfounded set *wrt* D and $\mathcal{I}_{D,\lambda-1}$. And so $p \in \mathcal{U}_{D,\lambda}$.

If λ is a limit ordinal, then if $p \in \mathcal{T}_{E,\lambda}$, there exists a least ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{E,\kappa}$. By the inductive hypothesis, we have $p \in \mathcal{T}_{D,\kappa}$. By monotonicity, $p \in \mathcal{T}_{D,\lambda}$. Similarly, if $p \in \mathcal{U}_{E,\lambda}$, there exists a least ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{E,\kappa}$. By the inductive hypothesis, we have $p \in \mathcal{U}_{D,\kappa}$. By monotonicity, $p \in \mathcal{U}_{D,\lambda}$. □

Note that D and E contain precisely the same number of rules. It is merely that certain rules have been moved from the set of defeaters to the set of defeasible rules. It is true that E contains potentially exponentially more conflict sets than D (for each conflict set $c \in C_D$, potentially $2^{|c|}$ new sets are generated). However, the conflict sets are related in such a fashion that it appears no more effort is required to evaluate the $2^{|c|}$ than to evaluate the original set c. *E.g.*, let S be the set of conflict sets created from basis c. If $q \in c$, then q appears in half of the members of S, and q^* appears in the other half. Since $|R_D[q]| = |R_E[q] \cup R_E[q^*]|$, the same number of rules need to be evaluated to evaluate c as to evaluate all of S.

CHAPTER 6

TWO MORE LOGICS

In previous chapters it was shown that BDL and its variants cannot identify indirect conflicts or detect circular reasoning, and also that ADL and NDL do not allow reinstatement; this latter behavior occurs even in unprioritized theories. In this chapter we present a prioritized logic called MDL which is intended to overcome these difficulties. However, MDL deviates considerably from ADL and NDL in how it defines conflict sets. Specifically, it defines conflict sets as sets of *rules* rather than literals. As this is so, we first describe a simpler logic called SDL which is structurally closer to ADL and NDL but which allows reinstatement. Like ADL and NDL, it incorporates extended conflict sets and failure-by-looping. However, it does not allow priorities among rules; in that sense it is simpler. While the operators defined for SDL propagate ambiguity, we show in later sections how they can be used to block ambiguity.

6.1 SDL: AN UNPRIORITIZED LOGIC

While the elimination of priorities limits the expressiveness of the logic, SDL is not without its virtues. First, in lacking priorities, the logic is easier to understand. From a practical standpoint, this is rather important. The two dominant semantics in logic programming—the well–founded semantics and the stable model/answer-set semantics—are unprioritized and also are among the simplest semantics to understand, and it is tempting to think that their popularity is due in part to their simplicity.

Second, under the translation of defeasible theories into logic programs, it turns out that the consequences of SDL correspond to the well-founded semantics. As this is so, the consequences under SDL are also the same as those obtainable using the operator α defined in Chapter 4. The

correspondence with the WFS is perhaps somewhat surprising. The research for this dissertation began with a suspicion that NDL might be similar to the well-founded semantics. It was found that a simple alteration to NDL formed an ambiguity propagating defeasible logic (ADL) that corresponds under certain restrictions to the WFS. However, it was also seen that NDL and ADL do not allow reinstatement. Now, in modifying the NDL/ADL proof–systems to incorporate reinstatement, we find ourselves even closer to the well-founded semantics.

In the following sections we describe how the operators U_D, T_D, and W_D defined in Chapter 4 can be modified to form SDL. Since many theorems stated here have proofs that proceed in much the same fashion as those for the corresponding theorems of ADL, the proofs have been relegated to Appendix D.

6.2 SEMANTICS

Definition 6.1. ***(Unfounded Sets in SDL)*** *A set S is* unfounded *in SDL with respect to D and an interpretation $\mathcal{I}$ iff for all literals $p \in S$:*

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U} \cup S) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

(2.1) $body(r) \cap (\mathcal{U} \cup S) \neq \varnothing$, or
(2.2) there is a $c \in C[p]$ such that for each $q \in c - \{p\}$, $q \in \mathcal{T}$.

There are two points to make about the above definition. The first is that it is simpler than the corresponding definition for ADL (the precedence relation is absent) but it places a stronger requirement on unfoundedness. Particularly, in order for a literal p to be unfounded, either every rule for it must have an unfounded literal in its body or else there is a conflict set c containing p for which all other literals of c are *already known to be well-founded*. Previously, we merely needed to show that each literal in $c - \{p\}$ had a *supported* rule s superior to r. This is stronger in the sense that it's possible in ADL for a defeated rule s to defeat r. The second point is that the above notion of unfounded set is really just the definition of unfounded set used in the WFS recast in the language of defeasible logic.

Like unfounded sets in ADL and NDL, unfounded sets for SDL are closed under union and so there exists a greatest unfounded set *wrt* D and $\mathcal{I}$.

Definition 6.2. $U_D(\mathcal{I}) = \bigcup\{S|\ S$ *is an unfounded set* wrt *to* D *and* $\mathcal{I}\}$.

The immediate consequence operator T_D for SDL is similarly simplified. Examining its definition below, one sees that one can detach the head from a rule r only if it is supported and strict or else if all conflict sets for $head(r)$ contain another member q *already known to be unfounded*. This is different than the corresponding operators for ADL and NDL, where (in the absence of priorities) one needed to show all rules for q failed to support q. One notes that a lack of support implies that q is unfounded, but not vice versa.

Definition 6.3 (Immediate Consequence in SDL)**.** $T_D(\mathcal{I}) = \{p|\ p$ *satisfies one of the conditions below*$\}$

(1) There exists an $r \in R_s[p]$ *such that* $body(r) \subseteq \mathcal{T}$.
(2) There exists a rule $r \in R_d[p]$ *such that* $body(r) \subseteq \mathcal{T}$ *and for each conflict set* $c \in C[p]$, $(c - \{p\}) \cap \mathcal{U} \neq \varnothing$.

As with ADL and NDL, the following propositions hold. The proofs appear in Appendix D; in the absence of priorities, these are fairly straightforward.

Lemma 6.4 (Unfounded sets are closed under union)**.** *If* S_0, S_2, ... *are SDL-unfounded sets* wrt *to defeasible theory* D *and interpretation* $\mathcal{I}$, *then* $\bigcup_{i=0}^{\infty} S_i$ *is SDL-unfounded* wrt *to* D *and* $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$.

Lemma 6.5 (T and U are monotone)**.** *: If* $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$, *then* $T_D(\mathcal{I}_1) \subseteq T_D(\mathcal{I}_2)$, *and* $U_D(\mathcal{I}_1) \subseteq U_D(\mathcal{I}_2)$.

T_D and U_D are again composed to form the monotone operator W_D. The well-founded model is defined as it is for ADL and NDL.

Definition 6.6 (Well-Founded Operator for SDL)**.** $W_D(\mathcal{I}) = \langle T_D(\mathcal{I}), U_D(\mathcal{I}) \rangle$.

W_D defines a sequence of interpretations $(\mathcal{I}_{D,0}, \mathcal{I}_{D,1}, \ldots)$; the well-founded model is defined as the least fixpoint of W_D.

Definition 6.7 (The Well-Founded Model under SDL)**.** *Let D be an unprioritized defeasible theory. The* well-founded model *of D under SDL, written $wfm_{SDL}(D)$, is defined to be $lfp(W_D)$.*

Proofs of the below appear in Appendix D.

Lemma 6.8 ($\mathcal{I}$ is monotone nondecreasing)**.** *For any $\alpha \geq 0$, if $p \in \mathcal{T}_\alpha$, then $p \in \mathcal{T}_{\alpha+1}$, and if $p \in \mathcal{U}_\alpha$, then $p \in \mathcal{U}_{\alpha+1}$.*

Lemma 6.9 (Each $\mathcal{I}_\alpha \in \mathcal{I}$ is coherent)**.** *For any $\alpha \geq 0$, $\mathcal{T}_\alpha \cap \mathcal{U}_\alpha = \varnothing$.*

6.2.1 EXAMPLES

SDL yields the correct results in many cases where ADL and NDL do not.

Example 6.10. *(3.23)*

(1) $\{\} \rightarrow married(chris)$
(2) $\{\} \Rightarrow \neg married(chris)$
(3) $\{\} \Rightarrow husband(chris)$
(4) $\{husband(chris)\} \rightarrow married(chris)$

Recall that $husband(chris)$ was refuted in NDL and undetermined in ADL. Under SDL we obtain the following sequence of interpretations:

0. $\langle \varnothing, \varnothing \rangle$
1. $\langle \{married(chris)\}, \{\neg husband(chris)\} \rangle$
2. $\langle \{married(chris)\}, \{\neg husband(chris), \neg married(chris)\} \rangle$
3. $\langle \{married(chris), husband(chris)\}, \{\neg husband(chris), \neg married(chris)\} \rangle$
4. $\langle \{married(chris), husband(chris)\}, \{\neg husband(chris), \neg married(chris)\} \rangle$

Rule r_2 conflicts with both r_1 and r_3. Starting with knowing nothing, since $married(chris)$ is supported by an empty strict rule, clearly $married(chris) \in T_D(\mathcal{I}_0) = \mathcal{T}_1$. However, since $\mathcal{T}_0 = \varnothing$, no set is unfounded *wrt* D and $\mathcal{I}_0$. However, since $married(chris) \in \mathcal{T}_1$ and $\neg married(chris)$ is supported only by a defeasible rule, $\neg married(chris) \in \mathcal{U}_2$. The only rule thus in conflict with r_3 has a head unfounded in $\mathcal{I}_2$. Thus $husband(chris) \in T_D(\mathcal{I}_2) = \mathcal{T}_3$.

Example 6.11. *(Example 1.3)*

(1) $\{\} \rightarrow livesAlone(joe)$
(2) $\{livesAlone(joe)\} \Rightarrow bachelor(joe)$
(3) $\{bachelor(joe)\} \rightarrow \neg married(joe).$
(4) $\{\} \rightarrow married(joe).$

We have the strict rule r_4 conflicting with the defeasible rule r_2, and so even though r_2 is supported we cannot conclude $bachelor(joe)$. This is what we want: Joe lives alone, and even though most men who live alone are bachelors, Joe is married. The results, which agree with NDL and ADL, are shown below. Recall that BDL also concludes that Joe is a bachelor.

$\langle \varnothing, \varnothing \rangle$

$\langle \{lvsAlne(joe), married(joe)\}, \{\neg lvsAlne(joe), \neg bachelor(joe)\} \rangle$

$\langle \{lvsAlne(joe), married(joe)\}, \{\neg lvsAlne(joe), \neg bachelor(joe), bachelor(joe)\} \rangle$

$\langle \{lvsAlne(joe), married(joe)\}, \{\neg lvsAlne(joe), \neg bachelor(joe), bachelor(joe), \neg married(joe)\} \rangle$

6.3 A Proof System

The semantics for SDL, similar to that for ADL and NDL, suggests that a proof system similar to those would serve for SDL.

Definition 6.12 (Proof in SDL)**.** *τ is a defeasible proof in SDL iff τ is an argument tree for D, and for each node n of τ, one of the following obtains:*

1. *n is labeled $+\delta p$ and either:*
   a. *there is rule $r \in R_s[p]$ such that $body(r)$ succeeds at n, or*
   b. *there is a rule $r \in R_d[p]$ such that*
      i. *$body(r)$ succeeds at n, and*
      ii. *for all $c \in C[p]$, $c - \{p\}$ fails at n.*
2. *n is labeled $-\delta p$ and:*
   a. *for all rules $r \in R_s[p]$, $body(r)$ fails at n, and*
   b. *for all rules $r \in R_d[p]$, either*
      i. *$body(r)$ fails at n, or*
      ii. *there is a $c \in C[p]$ such that $c - \{p\}$ succeeds at n.*

3. *n is labeled $-\delta p$ and has an ancestor m in τ labeled $-\delta p$, and all nodes between n and m are negative defeasible assertions.*

This proof procedure is sound and for finite grounded components complete *wrt* to the semantics of SDL. The definitions and lemmas for finite grounded components from Chapter 4 may be applied without change to SDL.

Theorem 6.13 (Soundness)**.** *Let $D = \langle R, C, \varnothing \rangle$ be a defeasible theory and $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, ..., the sequence of interpretations created by iterating W_D from $\langle \varnothing, \varnothing \rangle$.*

(1) If $D \mid\sim_{SDL} p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_{\mathcal{X}}$.
(2) If $D \sim\mid_{SDL} p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{U}_{\mathcal{X}}$.

Theorem 6.14 (Completeness for finite grounded components)**.** *Let $D = \langle R, C, \varnothing \rangle$ be a defeasible theory and $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, ..., the sequence of interpretations created by iterating W_D from $\langle \varnothing, \varnothing \rangle$. For any $\alpha \geq 0$, If $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then*

(1) if $p \in \mathcal{T}_{\mathcal{X}}$, then $D \mid\sim_{SDL} p$, and
(2) if $p \in \mathcal{U}_{\mathcal{X}}$, then $D \sim\mid_{SDL} p$.

6.4 Correspondence with the WFS

Returning to the topic of defeasible logic's relationship to logic programming, under the translation scheme presented in the previous chapter, the conclusions of a given defeasible theory under SDL correspond to the well-founded model of its logic program translation. Unlike the case with ADL, the correspondence holds even when the defeasible theory in question uses extended conflict sets (the restriction that $C[p]$ be finite for each p is still needed, however). The correspondence is even more extreme, however: Given D and its logic program translation Π_D, W_D and W_{Π_D} yield precisely the same sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots$. Since the proofs of this are somewhat different (and shorter) than those involving ADL, we include them here.

Theorem 6.15 (Soundness of SDL under the translation)**.** *Let $D = \langle R, C, \varnothing\rangle$ such that $R_u = \varnothing$ and for all $p \in Lit_D$ $C[p]$ is finite. Let Π be D's logic program translation, and let $\mathcal{I} = \langle \mathcal{T}, \mathcal{U}\rangle$ be an interpretation. If $p \in T_D(\mathcal{I})$, then $p \in T_\Pi(\mathcal{I})$. If $p \in U_D(\mathcal{I})$, then $p \in U_\Pi(\mathcal{I})$.*

Proof. Suppose $p \in T_D(\mathcal{I})$. Then by definition of T_D, either 1) there exists a strict rule $r \in R_{D,s}[p]$ such that $body(r) \subseteq \mathcal{T}$, or 2) there exists a defeasible rule $r \in R_{D,d}[p]$ such that $body(r) \subseteq \mathcal{T}$ and for all conflict sets $c \in C[p]$ there exists a $q \in c - \{p\}$ such that $q \in \mathcal{U}$. If 1 holds, then since r is strict, r appears unchanged in program Π. Thus, by definition of T_Π, $p \in T_\Pi(\mathcal{I})$.

Suppose 2 holds. Recall that, if $C[p]$ contains m conflict sets, then $r : A \Rightarrow p$ is translated into a set of logic program rules $r_1, r_2, \ldots$ of the form

$$p \ :- \ not\ q_1,\ not\ q_2, \ldots,\ not\ q_m, A.$$

where the i^{th} q is an element of $c_i \in C[p]$. Given that each conflict set has an element in $\mathcal{U}$, there is then a particular logic program rule $r_j \in trans(r)$ such that $body(r_j)^+ \subseteq \mathcal{T}$ and $body(r_j)^- \subseteq \mathcal{U}$. By definition of T_Π, $p \in T_\Pi(\mathcal{I})$.

Now Suppose $q \in U_D(\mathcal{I})$ for any arbitrary literal q. Then for each $r \in R_s[q]$, $body(r) \cap (\mathcal{U} \cup U_D(\mathcal{I})) \neq \varnothing$, and for all $r \in R_d[q]$, either $body(r) \cap (\mathcal{U} \cup U_D(\mathcal{I})) \neq \varnothing$, or there exists a $c \in C[q]$ such that $c - \{q\} \subseteq \mathcal{T}$. If $r \in R_s[q]$ then $r \in \Pi$. If $r \in R_d[q]$, then given the manner in which $trans(r)$ is defined, for each $r' \in trans(r)$, either $body(r')^+ \cap (\mathcal{U} \cup U_D(\mathcal{I})) \neq \varnothing$ or else there exists a $not\ a \in body(r')$ such that $a \in \mathcal{T}$. Generalizing on r' and r and then q, by definition of unfounded set for logic programs, $U_D(\mathcal{I})$ is unfounded *wrt* Π and $\mathcal{I}$. As such $p \in U_\Pi(\mathcal{I})$. □

The following lemma is used in the proof of completeness.

Lemma 6.16. *Let $D = \langle R, C, \prec\rangle$ be a defeasible theory such that $C[q]$ is finite for each $q \in Lit_D$. Let Π_D be the logic program translation of D. Let $\mathcal{T} \cup \{p\} \subseteq Lit_D$ and $C[p] = \{c_1, c_2, \ldots, c_m\}$. Let $r \in R_d[p]$. If for each $r' \in trans(r)$ there exists a $not\ b \in body(r')$ such that $b \in (c_i - \{p\}) \cap \mathcal{T}$ for some $i \leq m$, then there exists a $j \leq m$ such that $c_j - \{p\} \subseteq \mathcal{T}$.*

Proof. Suppose no such conflict set exists. Then for each $c_i \in C[p]$, there exists a $q \in c_i$ such that $q \notin \mathcal{T}$. Let $Q = \{q_1, q_2, \ldots, q_m\}$ be the set of such q's. Since one such q is included from each

$c \in C[p]$, it follows that $Q \in Prod(C[p])$. Since $r \in R_d$, there exists a $t \in trans(r)$ such that $Q = body(t)^-$. As such, for each $b \in body(t)^-$, $b \notin \mathcal{T}$. However, this contradicts our assumption that for each $r' \in trans(r)$ there exists a $not\ b \in body(r')$ such that $b \in (c_i - \{p\}) \cap \mathcal{T}$ for some $i \leq m$. And so it must be that there is a conflict set $c_j \in C[p]$ such that $c_j - \{p\} \subseteq \mathcal{T}$. □

Theorem 6.17 (Completeness under the translation)**.** *Let $D = \langle R, C, \varnothing \rangle$ be a defeasible theory such that $R_u = \varnothing$ and for all $p \in Lit_D$, $C[p]$ is finite. Let Π be its logic program translation. Let $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$. If $p \in T_\Pi(\mathcal{I})$, then $p \in T_D(\mathcal{I})$. If $p \in U_\Pi(\mathcal{I})$, then $p \in U_D(\mathcal{I})$.*

Proof. Suppose $p \in T_\Pi(\mathcal{I})$. Then by definition of T_Π, there exists an $r' \in trans(r)$ such that $body(r')^+ \subseteq \mathcal{T}$ and $body(r')^- \subseteq \mathcal{U}$. If $body(r')^- = \varnothing$, then $r' = r$ and $body(r) = body(r')^+$. If $body(r')^- \neq \varnothing$, then r is defeasible. However, since $body(r')^- \subseteq \mathcal{U}$, for each conflict set $c \in C[p]$ there exists a $q \in c - \{p\}$ such that $q \in \mathcal{U}$. Thus, either there exists a strict rule $r \in R_{D,s}[p]$ such that $body(r) \subseteq \mathcal{T}$, or else a defeasible rule $r \in R_{D,d}[p]$ such that $body(r) \subseteq \mathcal{T}$ and for all conflict sets $c \in C[p]$, there exists a $q \in c - \{p\}$ such that $q \in \mathcal{U}$. Either way, $p \in T_D(\mathcal{I})$ by definition of T_D.

Now suppose $p \in U_\Pi(\mathcal{I})$ and let q be any literal such that $q \in U_\Pi(\mathcal{I})$. Then by definition of U_Π, for all rules $r' \in \Pi$ with head q either $body(r')^+ \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$ or else $body(r')^- \cap \mathcal{T} \neq \varnothing$. Suppose r' corresponds to a strict rule $r \in R_s[q]$. Then $body(r')^- = \varnothing$ and hence $body(r) \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$. Suppose $r' \in trans(r)$ for a defeasible rule r. If $body(r')^+ \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$, then $body(r) \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$. Furthermore, since for each $s, t \in trans(r)$ it holds that $body(s)^+ = body(t)^+$, it follows that for each $s \in trans(r)$, $body(s) \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$. Suppose for each $s \in trans(r)$, $body(s)^+ \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) = \varnothing$. Then by definition of $U_\Pi(\mathcal{I})$, it must be the case that for each $s \in trans(r)$, $body(s)^- \cap \mathcal{T} \neq \varnothing$. By Lemma 6.16, there must be a conflict set $c \in C[q]$ such that $c - \{q\} \subseteq \mathcal{T}$. Thus, for each rule $r \in R_s[q]$, $body(r) \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$, and for each $r \in R_d[q]$, either $body(r) \cap (\mathcal{U} \cup U_\Pi(\mathcal{I})) \neq \varnothing$ or there exists a conflict set $c \in C[q]$ such that $c - \{q\} \subseteq \mathcal{T}$. Generalizing on q, $U_\Pi(\mathcal{I})$ is unfounded *wrt* to D and $\mathcal{I}$. And so $U_\Pi(\mathcal{I}) \subseteq U_D(\mathcal{I})$. As such $p \in U_D(\mathcal{I})$. □

And so, if $R_u = \prec = \varnothing$ and $C[p]$ is finite for all p, the well-founded model of the defeasible theory under SDL corresponds to the well-founded model of the logic program translation. Since it is known that the WFS satisfies Cut and Cautious monotony (see, for instance, [Dix94]) we can infer that SDL satisfies these properties as well. Also, given the correspondence between α_D and γ_{Π_D}, it follows that of δ_D can be used to compute the consequences of theories under SDL.

Theorem 6.18. *Let $D = \langle R, C, \varnothing \rangle$ be a defeasible theory such that $R_u = \varnothing$ and for all $p \in Lit_D$, $C[p]$ is finite. Let $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ be its well-founded model under SDL. Then $\mathcal{T} = lfp(\delta)$ and $\mathcal{U} = Lit_D - \alpha(\mathcal{T})$.*[1]

6.5 Semi-Normal Defeasible Theories

Extended conflict sets have been proposed as a mechanism for handling, in a computationally tractable manner, indirect conflicts in theories. However, strategies for generating conflict sets have not been discussed, and in truth closing conflict sets under strict rules leads to exponentially many conflict sets. In this section we discuss defeasible theories with minimal conflict sets, no defeaters, an empty precedence relation, and strict rules closed under transposition. We will call such theories *semi-normal* (SN). Semi–normal theories have many desirable properties. If the strict portion of the theory is consistent, then they correspond exactly to *semi-normal extended (SNE) logic programs*, a class of logic program guaranteed to have a consistent well-founded model [Cam06]. The translation of semi–normal theory into its corresponding logic program is modular and can be performed efficiently. Below we define SNE logic programs and SN defeasible theories, and prove that SN theories have consistent well-founded models. Observe that the terminology used in [Cam06] differs from that discussed in Chapter 2. Here a semi-normal extended program requires that strict rules be closed under transposition.

Definition 6.19 (Strict and defeasible logic program rules)**.** *A logic program rule r is* defeasible *if it contains an occurrence of a default literal. Otherwise, r is* strict*.*

[1] Note that we have not included defeaters. We suspect that the correspondence holds when defeaters are used as well. However, we have not yet proven this. However, as was shown in the previous chapter, defeaters can be safely removed from the theory.

Definition 6.20 (Transposition of Rules)**.** *Let* $r : b \ :- \ a_1, a_2, \ldots a_n$ *be a strict rule. Define* $transposed(r)$ *as*

$$transposed(r) = \{r' |\ r' = \neg a_i \ :- a_1, a_2, \ldots, a_{i-1}, \neg b, a_{i+1}, \ldots, a_n \textit{ for some } 1 \leq i \leq n\}$$

A set S *of strict rules is closed under transposition iff for all* $r \in S$ *we have* $trans(r) \subseteq S$*.*

Transposition can be defined analogously for strict rules in defeasible theories.

Definition 6.21 (Consistent Sets)**.** *A set of literals* X *is consistent* iff *for any atom* p*,* $\{p, \neg p\} \nsubseteq X$*. A set of strict rules* S *is consistent* iff $Cl(S)$ *is consistent.*

Definition 6.22 (Semi-Normal Extended Logic Programs [Cam06])**.** *Let* Π *be an extended logic program,* $strict(\Pi)$ *the set of strict rules of* Π*, and* $defeasible(\Pi)$ *the set of defeasible rules of* Π*.*

(1) A rule $r \in defeasible(\Pi)$ *is* semi-normal iff $not\ \neg head(r) \in body(r)$*.*
(2) Program Π *is semi-normal* iff

- *(2.1)* $strict(\Pi)$ *is consistent,*
- *(2.2)* $strict(\Pi)$ *is closed under transposition, and*
- *(2.3) each* $r \in defeasible(\Pi)$ *is semi-normal.*

Recall that a 4-valued interpretation X of a program Π as any subset of Lit_Π. The well-founded model is presented in [Cam06] as $lfp(\gamma_\Pi^2)$, where $\gamma_\Pi(S) = Cl(\Pi^S)$. Note that given the relationship between W_Π and γ_Π, $lfp(\gamma_\Pi^2)$ corresponds to $\mathcal{T}$ in the $lfp(W_\Pi) = \langle \mathcal{T}, \mathcal{U} \rangle$.

Theorem 6.23 (SNE Programs are consistent [Cam06])**.** *Let* Π *be a semi-normal extended logic program. Then the well-founded model of* Π *is consistent.*

Given the above notion of a SNE program, we define semi-normal defeasible theories.

Definition 6.24 (Semi-Normal Defeasible Theories)**.** *Let* $D = \langle R, C, \prec \rangle$ *be a defeasible theory.* D *is* semi–normal *iff* C *is minimal,* $\prec = \varnothing$*,* $R_u = \varnothing$*, and* R_s *is closed under transposition.*

Theorem 6.25 (SN Defeasible Theories are Consistent)**.** *Let* D *be a semi–normal defeasible theory and* Π *its logic program translation. Let* R_s *be the strict rules of* D*. If* R_s *is consistent, then* $wfm(D)$ *under SDL is consistent.*

Proof. Let $wfm_{SDL}(D)$ be the well-founded model of D according to SDL and M_Π the well-founded model of Π. Since D contains no defeaters and the precedence relation is empty, then by Theorems 6.15 and 6.17 (soundness and completeness under the translation) we have $wfm_{SDL}(D) = M_\Pi$. Suppose R_s is consistent. By definition of semi–normal theory, R_s is closed under transposition. Since there is a 1-1 correspondence between the strict rules of D and the strict rules of Π, it follows that $strict(\Pi)$ is both consistent and closed under transposition. Since the conflict sets of D are minimal, there is also a 1-1 correspondence between the defeasible rules of D and Π, and if $r \in R_d$, then its counterpart $r' \in \Pi$ is in semi-normal form by definition of the usual translation scheme. Thus Π is a semi-normal program. As such, M_Π is consistent. Since $M_\Pi = wfm_{SDL}(D)$, $wfm_{SDL}(D)$ is consistent. □

6.5.1 EXAMPLES

The Semi-Normal form of Example 3.23 is shown below. Since conflict sets are minimal, it is not the case that $\{husband(chris), \neg married(chris)\}$ conflict.

Example 6.26. *(3.23)*

(1) $\{\} \rightarrow married(chris)$
(2) $\{\} \Rightarrow \neg married(chris)$
(3) $\{\} \Rightarrow husband(chris)$
(4) $\{husband(chris)\} \rightarrow married(chris)$
(5) $\{\neg married(chris)\} \rightarrow \neg husband(chris)$

The sequence of interpretations produced by the semi-normal theory match those obtained with the original.

0. $\langle \varnothing, \varnothing \rangle$
1. $\langle \{married(chris)\}, \varnothing \rangle$
2. $\langle \{married(chris)\}, \{\neg married(chris)\} \rangle$
3. $\langle \{married(chris)\}, \{\neg married(chris), \neg husband(chris)\} \rangle$
4. $\langle \{married(chris), husband(chris)\}, \{\neg married(chris), \neg husband(chris)\} \rangle$

Example 6.27.

(1) $\{\} \rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $\{p\} \rightarrow r$
(4) $\{q\} \rightarrow \neg r$
(5) $\{\neg q\} \rightarrow s$.

If conflict sets are closed in the first theory, the sets are: $\{p, \neg r\}$, $\{q, r\}$ $\{\neg q, \neg s\}$, $\{p, q\}$, $\{r, \neg s\}$, and $\{p, \neg s\}$. Under SDL, the well-founded model is $\langle\{p, r\}, \{q, \neg q, \neg r, s\}\rangle$. If we include literals not mentioned explicitly in the theory, the model is $\langle\{p, r\}, \{\neg p, q, \neg q, \neg r, s, \neg s\}\rangle$.

The semi-normal form of the theory appears below.

(1) $\{\} \rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $p \rightarrow r$
(4) $\neg r \rightarrow \neg p$
(5) $q \rightarrow \neg r$
(6) $r \rightarrow \neg q$
(7) $\neg q \rightarrow s$
(8) $\neg s \rightarrow q$

The model under SDL is $\langle\{p, r, \neg q, s\}, \{\neg p, \neg r, q, \neg s\}\rangle$.

6.6 Blocking Ambiguity

Example 6.28.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow \neg p$

In the above simple theory, r_1 and r_2 conflict, and in the absence of priorities, there is no way to choose between them. p and $\neg p$ are ambiguous. In SDL, it is impossible to either prove or refute either p or $\neg p$, and this potentially affects the status of rules depending upon them. SDL is ambiguity propagating.

In general, the well-founded model of a defeasible theory might contain many ambiguous literals. However, some of these literals will in some sense be *primitively ambiguous—e.g.*, supported rules for p and $\neg p$ exist but neither is preferred to the other. Other literals will merely be

derivatively ambiguous—ambiguous because they depend on other ambiguous literals. Recall from Chapter 1 that ambiguity blocking behavior could be achieved in inheritance networks by deleting the nodes at which a conflict initially occurs. By performing roughly the same action on rules of defeasible theories, the operators of SDL can be used to define an ambiguity blocking relation. That is, after calculating the well-founded model of D under SDL, we determine the set of primitively ambiguous literals defined by this model and then delete all rules for these literals. Once that is done, we compute a new model for the modified theory. This process can be continued until a fixpoint is reached.

A definition of primitive ambiguity is shown below. Note however that the iterative strategy described above works for *any* chosen set of ambiguous literals and is not, strictly speaking, related to any notion of primitiveness. Because of this, the strategy provides considerable flexibility. One might choose to block ambiguity for some literals and not others. This is an important point to make, for if intuitions can clash over whether ambiguity blocking is reasonable *as a general policy*, intuitions might also clash over whether ambiguity should propagated for a particular literal in a particular example.

Definition 6.29 (Ambiguous Literal)**.** *Let D be a defeasible theory and $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ an interpretation. A literal p is ambiguous* wrt *to D and $\mathcal{I}$ iff $p \notin (\mathcal{T} \cup \mathcal{U})$.*

Definition 6.30 (Primitive Ambiguity in SDL)**.** *A literal p is primitively ambiguous* wrt *a defeasible theory D and interpretation $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$ iff p is ambiguous* wrt *D and $\mathcal{I}$ and there exists a conflict set $c \in C[p]$ that may be partitioned into sets A and B such that*

(1) for all $q \in B$, $q \in \mathcal{T}$, and
(2) $p \in A$,
(3) for all $q \in A$,

- *(3.1) q is ambiguous* wrt *D and $\mathcal{I}$, and*
- *(3.2) for all $r \in R_s[q]$, $body(r) \cap \mathcal{U} \neq \varnothing$,*
- *(3.3) there exists an $r \in R_{du}[q]$ such that $body(r) \subseteq \mathcal{T}$.*

Given that *some* set of privileged literals can be chosen from an interpretation and theory, we can use this set and the SDL operators already presented to create an ambiguity blocking relation.

Definition 6.31 (The Residual of D)**.** *Let D be a defeasible theory and $\mathcal{I}$ an interpretation. Let A be a set of ambiguous literals obtained from $\mathcal{I}$. The residual of D* wrt *to $\mathcal{I}$ and A, written $D^{\mathcal{I},A}$, is the theory obtained by performing the following:*

(1) Delete from D all rules r such that $\{head(r)\} \cup body(r) \in (\mathcal{U}_{\mathcal{I}} \cup A)$.
(2) Delete from $\prec_D$ any element in which a deleted rule occurs.

The idea behind the first step above is that rules with heads or bodies in $\mathcal{U}$ or A are known never to be satisfied and so can be deleted. In the second step, since these rules no longer appear in the theory, they serve no purpose in the precedence relation.

Given the notion of residual theories, we can define a new operation which maps interpretations to interpretations. Since we do not wish to commit to ourselves entirely to the notion of primitive ambiguity provided, we define a family of operators, each based on a particular function which selects some set of ambiguous literals.

Definition 6.32 (Selection Function)**.** *Let D be a defeasible theory, $\mathcal{I}$, an interpretation, and X the set of ambiguous literals* wrt *D and $\mathcal{I}$. A* selection function *f is any function which maps D and $\mathcal{I}$ to a subset of X.*

Definition 6.33. *Let D be a defeasible theory, $\mathcal{I}$ an interpretation, and f a selection function.* $\Psi_{D,f}(\mathcal{I}) = wfm(D^{\mathcal{I},f(D,\mathcal{I})})$.

Given the above, we can define a sequence of interpretations $J_0, J_1, \ldots$:

$$
\begin{aligned}
&J_0 = \langle \varnothing, \varnothing \rangle \\
&J_{\alpha+1} = \Psi_{D,f}(J_\alpha) && \text{(for successor ordinals)} \\
&J_\alpha = lub(\{J_\beta | \beta < \alpha\}) && \text{(for limit ordinals.)}
\end{aligned}
$$

We shall use A_{J_α} to refer to $f(D, J_\alpha)$ (*i.e.*, the set of ambiguous literals chosen from J_α by f). As the below lemma shows, the sequence of J's, like the sequence $\mathcal{I}$, is monotonically non-decreasing. Since $J_1 = \mathcal{I}_{D,WF}$, by monotonicity we have for each $\alpha > 0$, $J_1 \sqsubseteq J_\alpha$, and so all of the results of SDL are contained in subsequent interpretations. In other words, the above process preserves but perhaps extends the consequences obtainable under SDL.

Lemma 6.34 (J is monotonically nondecreasing)**.** *Let D be an unprioritized defeasible theory and J_0, J_1, ... the sequence of interpretations defined by iterating $\Psi_{D,f}$ from the empty set. For all λ, if $\kappa < \lambda$, then $J_\kappa \sqsubseteq J_\lambda$.*

Proof. The claim trivially holds for $\lambda = 0$. Suppose it holds for all $\alpha < \lambda$. We will show for any $\kappa > 0$, if $\kappa < \lambda$, then $J_\kappa \sqsubseteq J_\lambda$. This obviously holds for $\kappa = 0$. Suppose it holds for all $\beta < \kappa$. If λ is a limit ordinal, then since $\kappa < \lambda$ we have $J_\kappa \sqsubseteq J_\lambda$ by definition of J_λ. Suppose λ is a successor ordinal. Either $\kappa = \lambda - 1$ and so $J_\kappa = J_{\lambda-1}$ or else by the inductive hypothesis $J_\kappa \sqsubseteq J_{\lambda-1}$.

Observe that for any $p \in \mathcal{U}_{J_\kappa}$ or any $p \in \mathcal{T}_{J_\kappa}$, if κ is a limit ordinal, then there exists a $\beta < \kappa$ such that $p \in \mathcal{U}_{J_\beta}$ ($p \in \mathcal{T}_{J_\beta}$), and so by hypothesis $p \in \mathcal{U}_{J_\lambda}$ ($p \in \mathcal{T}_{J_\lambda}$). Suppose κ to be a successor ordinal.

(Case 1) Suppose $p \in \mathcal{U}_{J_\kappa}$. Then $p \in \mathcal{U}_{J_{\lambda-1}}$. Since by definition of $D^{J_\lambda, A_{J_\lambda}}$ every rule for p is deleted in $D^{J_\lambda, A_{J_\lambda}}$, it is clear that $p \in \mathcal{U}_{J_\lambda}$.

(Case 2) Suppose $p \in \mathcal{T}_{J_\kappa}$. We will induct on the sequence of interpretations

(1) $J_{\kappa,0} = \langle \varnothing, \varnothing \rangle$
(2) $J_{\kappa,1} = W_{D^{J_{\kappa-1}, A_{J_{\kappa-1}}}}(J_{\kappa,0})$
(3) $J_{\kappa,2} = W_{D^{J_{\kappa-1}, A_{J_{\kappa-1}}}}(J_{\kappa,1})$
(4) ...

used to define J_κ (*i.e.*, the sequence that defines $wfm(D^{J_{\kappa-1}, A_{J_{\kappa-1}}})$), showing that for all $\eta \geq 0$, if $p \in J_{\kappa,\eta}$, then $p \in \mathcal{T}_{J_\lambda}$. Suppose the claim holds for all $\eta < \mu$ and let $p \in \mathcal{T}_{J_{\kappa,\mu}}$. Suppose μ is a successor ordinal. By definition of T_D either (2.1) there is an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_{J_\kappa,\mu-1}$, or else (2.2) there is an $r \in R_D[p]$ such that $body(r)\mathcal{T}_{J_\kappa,\mu-1}$ and for all conflict sets $c \in C[p]$ we have $(c - \{p\}) \cap \mathcal{U}_{J_\kappa,\mu-1} \neq \varnothing$. We'll treat each case in turn. Note that since $body(r) \subseteq \mathcal{T}_{J_{\kappa,\mu-1}}$, we have $p \in \mathcal{T}_{J_\kappa}$ and $body(r) \subseteq \mathcal{T}_{J_\kappa}$. By hypothesis, this means we have $p \in \mathcal{T}_{J_{\lambda-1}}$ and $body(r) \subseteq \mathcal{T}_{J_{\lambda-1}}$, and by coherence we have $(\{p\} \cup body(r)) \cap \mathcal{U}_{J_{\lambda-1}} = \varnothing$. Since $p \in \mathcal{T}_{J_{\lambda-1}}$, p cannot be ambiguous there. Since this is so and $(\{p\} \cup body(r)) \cap \mathcal{U}_{J_{\lambda-1}} = \varnothing$, rule r appears in $D^{J_{\lambda-1}, A_{J_{\lambda-1}}}$.

(Case 2.1) If $r \in R_s[p]$ and $body(r) \subseteq \mathcal{T}_{J_\kappa,\mu-1}$, then by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{J_\lambda}$. Since r is strict and $body(r) \subseteq \mathcal{T}_{J_\lambda}$ and $r \in D^{J_{\lambda-1}, A_{J_{\lambda-1}}}$, clearly $p \in \mathcal{T}_{J_\lambda}$.

(Case 2.2) Since there is an $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_{J_\kappa,\mu-1}$, by the inductive hypothesis we have $body(r) \subseteq \mathcal{T}_{J_\lambda}$. Since for all conflict sets $c \in C[p]$ we have $(c - \{p\}) \cap \mathcal{U}_{J_\kappa,\mu-1} \neq \varnothing$, by monotonicity of $\mathcal{U}$ it must be that for some $v \in (c - \{p\})$ we have $v \in \mathcal{U}_{J_\kappa}$. As in Case 1, no rule containing v in the head or body appears in $D^{J_{\lambda-1},A_{J_{\lambda-1}}}$. Generalizing on c, since r appears in $D^{J_{\lambda-1},A_{J_{\lambda-1}}}$ and $body(r) \subseteq \mathcal{T}_{J_\lambda}$ and each $c \in C[p]$ contains a member distinct from p with no rules in $D^{J_{\lambda-1},A_{J_{\lambda-1}}}$, it follows that $p \in \mathcal{T}_{J_\lambda}$.

If μ is a limit ordinal, then there is some successor ordinal $\eta < \mu$ such that $p \in \mathcal{T}_{J_{\kappa,\eta}}$. By hypothesis, $p \in \mathcal{T}_{J_\lambda}$. □

6.7 The Coherence Principle Revisited

Consider the below theory D (Assume the precedence relation is empty).

Example 6.35.

(1) $\{\} \rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $\{\} \Rightarrow r$
(4) $\{q, r\} \rightarrow \neg p$
(5) $\{\neg p\} \Rightarrow s$
(6) $\{\} \Rightarrow \neg s$

The extended conflict sets are SDL are $\{p, \neg p\}$, $\{q, \neg q\}$, $\{r, \neg r\}$, $\{s, \neg s\}$, and $\{p, q, r\}$. $M = \langle\{p\}, \{\neg q, \neg r\}\rangle$ is the well-founded model of D according to SDL. The literals q and r are both primitively ambiguous—there is no way to choose one over the other—which means that both $\neg p$ and s are derivatively ambiguous ($\neg p$ does not satisfy the definition of primitive ambiguity, however). The result is somewhat puzzling, however, in that p is defeasibly proven but $\neg p$ is ambiguous.

Recall that in the context of extended logic programs, Alferes and Pereira recommend imposing a "coherence restriction" upon programs [PA92], [ADP95] (again, for any literal p, if $p \in \mathcal{T}$ then $\neg p \in \mathcal{U}$). We can incorporate this requirement into SDL by precisely the mechanism proposed in

the last section for other ambiguous literals.[2] If after computing the fixpoint M of W_D, we find $p \in \mathcal{T}_M$ but $\neg p \notin (\mathcal{U}_M \cup \mathcal{T}_M)$, we can remove all rules for $\neg p$ (in the present example, the rule deleted is strict) and then compute the fixpoint of the residual theory. Note, however, that in all versions of defeasible logic, it is possible for both p and $\neg p$ to be derivable. Because of this, we do not wish to completely adhere to the coherence principle. We feel it should only be enforced if $p \in \mathcal{T}_M$ but $\neg p \notin (\mathcal{U}_M \cup \mathcal{T}_M)$.

In the above example, the change is significant, for it means that s becomes unfounded. The well-founded model of D under SDL with the coherence restriction is: $\langle\{p, \neg s\}, \{\neg p, s, \neg q, \neg r\}\rangle$. This indeed is the desired result. Given that $\{p, q, r\}$ is a conflict set and p is a fact of the theory, q and r can't both hold. Regardless of which one doesn't, the support for $\neg p$, and indirectly for s, collapses. Thus no supported rule conflicts with r_6, and so $\neg s$ is derivable.

6.8 MDL: Reincorporating Priorities

We now define the logic MDL, an ambiguity propagating logic which both allows rule reinstatement and incorporates priorities. MDL redefines conflict sets to be sets of non–strict rules and so does not extend any of the previously defined logics. The altered definition is shown below. Intuitively, a set of defeasible rules S is a conflict set if the heads of rules in S can be used together with the strict rules of D to conclude a contradiction. This notion is formalized below.

Definition 6.36 (Defeasible Set of Support)**.** *Let D be a defeasible theory, p a literal of D, and $X \subseteq R_{D,d}$.*

(1) *X is a* defeasible set of support *for p if*

  (1.1) *for all $r_1, r_2 \in X$, if $head(r_1) = head(r_2)$ then $r_1 = r_2$, and*
  (1.2) *$R_s \cup \{\{\} \rightarrow head(r) |\ r \in X\} \vdash p$ and for all $T \subset X$, $R_s \cup \{\{\} \rightarrow head(r) |\ r \in T\} \nvdash p$*

(2) *$Supp(p)$ is the set $\{X |\ X$ is a defeasible set of support for $p\}$*

[2] Since prioritized theories are allowed in the other logics, and since deleting rules will likely affect the impact of the priority relation, it seems doubtful that the technique discussed here will work "properly" for those logics.

Definition 6.37 (Basic Conflict Sets in MDL)**.** *A set $S \subseteq (R_d \cup R_u)$ is a* basic conflict set iff

(1) for all $r_1, r_2 \in S$, if $head(r_1) = head(r_2)$ then $r_1 = r_2$, and
(2) there exists an atom p such that $R_s \cup \{\{\} \rightarrow head(r) |\ r \in S\} \vdash \{p, \neg p\}$ and for all $T \subset S$, $R_s \cup \{\{\} \rightarrow head(r) |\ r \in T\} \nvdash \{p, \neg p\}$

Definition 6.38 (Conflict Sets in MDL)**.** *A set $C_1 \subseteq (R_d \cup R_u)$ is a* conflict set *if it is basic or if C_2 is a conflict set and*

(1) $r \in C_2$ and
(2) $X \in Supp(head(r))$ and
(3) $C_1 = (C_2 - \{r\}) \cup X$.

'$\vdash$' above is here taken to mean that p is derivable via repeated applications of a monotone immediate consequence operator. In general, it will be very difficult to enumerate all the sets of the conflicting rules (just as it is difficult to determine conflicting sets of literals). This is precisely why conflict sets are explicitly given as part of the defeasible theory. It allows one to avoid under some circumstances computational intractability.

One might wonder why we have chosen to redefine conflict sets. The below example helps explain why.

Example 6.39.

(1) $\{\} \Rightarrow p$
(2) $\{q\} \rightarrow \neg p$
(3) $\{\} \Rightarrow q$

$3 \prec 1$

If we define conflict sets as sets of *literals* and close conflict sets under strict rules, then $\{p, q\}$ and $\{p, \neg p\}$ are both conflict sets. In NDL and ADL, the mere fact that rule 1 is supported is sufficient to defeat rule 3. Because of this, q and $\neg p$ are unfounded. However, to incorporate reinstatement, we required not only that rule 1 be supported but that it actually fire. That is, in order to refute q, we must first prove p. However, since $\{p, \neg p\}$ is a conflict set, the only way we can show p is by first showing that rule 2 fails. But this can be done only if q is refuted. There is thus

no way to get the process started. In MDL, the solution to the problem is to define conflict directly between rules. In the example, rule 1 conflicts, not with rule 2, but with the defeasible support for 2—namely, rule 3. Since rule 3 is inferior to rule 1 and rule 1 is supported, we are allowed to conclude p. And since rule 1 actually fires, rule 3 is defeated, and so q is refuted.

6.9 SEMANTICS

Definition 6.40 (Unfounded Sets in MDL)**.** *A set S is* unfounded *in MDL with respect to D and an interpretation $\mathcal{I}$ iff for all literals $p \in S$:*

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U} \cup S) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

- *(2.1) $body(r) \cap (\mathcal{U} \cup S) \neq \varnothing$, or*
- *(2.2) there is a $c \in C[r]$ such that for each $s \in c - \{r\}$,*
  - *(2.2.1) $\{head(s)\} \cup body(s) \subseteq \mathcal{T}$,*
  - *(2.2.2) $s \not\prec r$*

The above definition differs from that for SDL in that (1) conflict sets are defined over rules, and (2) every rule in $c - \{r\}$ not only must have a head that is in $\mathcal{T}$, but $body(s)$ must be in $\mathcal{T}$ and $s \not\prec r$ as well. The restriction seems at first somewhat odd, since if $head(s)$ is in $\mathcal{T}$ then presumably there is *some* rule with $head(s)$ that succeeds and that is at least as good r. However, this holds only if conflict sets are defined in a natural way such as above. If arbitrary conflict sets are used, or if the set of conflict sets is restricted due to limitations in computational resources, then the above condition will not be met in general. Indeed, if conflict sets are arbitrary, then the restriction that $body(s) \subseteq \mathcal{T}$ and $s \not\prec r$ is needed to ensure coherence.

Unfounded sets for MDL are closed under union and we use this fact to define $U_D(\mathcal{I})$.

Definition 6.41. $U_D(\mathcal{I}) = \bigcup\{S|\ S$ *is an unfounded set* wrt *to D and $\mathcal{I}$*$\}$.

The immediate consequence operator T_D for MDL appears below. Examining its definition, one sees that one can detach the head from a rule r only if $body(r)$ is well-founded and r is strict or if all conflict sets for r contain another member s such that $head(s)$ is already known to be unfounded,

or $body(s)$ itself is unfounded, or $s \prec r$. This is different than the corresponding operator for ADL and NDL. Indeed, this is how MDL incorporates reinstatement; the fact that $head(s)$ is unfounded serves as a witness that s has been defeated.

Definition 6.42. $T_D(\mathcal{I}) = \{p|$ *p satisfies one of the conditions below*$\}$

(1) There exists an $r \in R_s[p]$ *such that* $body(r) \subseteq \mathcal{T}$.
(2) There exists a rule $r \in R_d[p]$ *such that*

(2.1) $body(r) \subseteq \mathcal{T}$ *and*
(2.2) for each conflict set $c \in C[r]$*, there exists a* $s \in c - \{r\}$ *such that either*

(2.2.1) $(\{head(s)\} \cup body(s)) \cap \mathcal{U} \neq \varnothing$*, or*
(2.2.2) $s \prec r$.

Definition 6.43. $W_D(\mathcal{I}) = \langle T_D(\mathcal{I}), U_D(\mathcal{I})\rangle$.

Definition 6.44 (The Well-Founded Model for MDL)**.** *Let* D *be a defeasible theory. The* well-founded model *of* D *under MDL, written* $wfm_{MDL}(D)$*, is defined to be* $lfp(W_D)$.

As with the other logics, certain propositions must be proven in order to show that the well-founded model is well-defined. The proofs appear in Appendix E.

6.9.1 EXAMPLES

Consider again an example from previous chapters.

Example 6.45.

(1) $\{\} \rightarrow married(chris)$
(2) $\{\} \Rightarrow \neg married(chris)$
(3) $\{\} \Rightarrow husband(chris)$
(4) $\{husband(chris)\} \rightarrow married(chris)$

The only conflict set (basic or otherwise) according to MDL is $\{r_2\}$. The set $\{r_2, r_3\}$ would also derive a contradiction, but $\{r_2\} \subset \{r_2, r_3\}$ and so $\{r_2, r_3\}$ is not a valid conflict set. Without priorities, the sequence of interpretations is:

0. $\langle \varnothing, \varnothing \rangle$
1. $\langle \{married(chris), husband(chris)\}, \{\neg married(chris), \neg husband(chris)\} \rangle$

Since r_1 is strict and nothing conflicts with r_3 and both are supported in $\mathcal{T}_0$, both $married(chris) \in \mathcal{T}_1$ and $husband(chris) \in \mathcal{T}_1$. However, since $\{r_2\}$ is a conflict set and $\{r_2\} - r_2$ is empty, it follows that $\{\neg married(chris)\}$ is unfounded *wrt* D and $\mathcal{I}_0$. And so $\neg married(chris) \in \mathcal{U}_1$. Modifying the example so that r_1 is defeasible and by adding priorities, then different results are obtained.

Example 6.46.

(1) $\{\} \Rightarrow married(chris)$
(2) $\{\} \Rightarrow \neg married(chris)$
(3) $\{\} \Rightarrow husband(chris)$
(4) $\{husband(chris)\} \rightarrow married(chris)$

The conflict sets are now $\{r_1, r_2\}$ and $\{r_3, r_2\}$. Suppose $r_1 \prec r_2$ and $r_3 \prec r_2$. We then get the following.

0. $\langle \varnothing, \varnothing \rangle$
1. $\langle \{\neg married(chris)\}, \{\neg husband(chris)\} \rangle$
2. $\langle \{\neg married(chris)\}, \{married(chris), husband(chris), \neg husband(chris)\} \rangle$

Since r_2 is superior to both r_1 and r_3, and these latter rules are the only ones which conflict with r_2, then it is safe to conclude r_2. However, with $head(r_2) \in \mathcal{T}_1$ and $body(r_2) \subseteq \mathcal{T}_1$ and $r_2 \not\prec r_1$ and $r_2 \not\prec r_3$, $married(chris)$ and $husband(chris)$ become unfounded. Note, however, if the priorities are eliminated, then nothing can be concluded from the theory. Nor could anything be concluded if $r_1 \prec r_2$ but $r_3 \not\prec r_2$ (or vice versa).

Example 6.47. *(Example 1.3)*

(1) $\{\} \rightarrow lAlone(joe)$
(2) $\{lAlone(joe)\} \Rightarrow bachelor(joe)$
(3) $\{bachelor(joe)\} \rightarrow \neg married(joe)$.
(4) $\{\} \rightarrow married(joe)$.

$\{r_2\}$ is again a conflict set and so $bachelor(joe)$ is automatically unfounded. This is the desired result. The sequence of interpretations produced using MDL converges after only two iterations.

(1) $\langle \varnothing, \varnothing \rangle$
(2) $\langle \{lAlone(joe), married(joe)\}, \{\neg lAlone(joe), bachelor(joe), \neg bachelor(joe)\}\rangle$
(3) $\langle \{lAlone(joe), married(joe)\}, \{\neg lAlone(joe), bachelor(joe), \neg bachelor(joe)\neg married(joe)\}\rangle$

Example 6.48.

(1) $\{\} \Rightarrow p$
(2) $\{\} \Rightarrow q$
(3) $\{\} \rightarrow q$
(4) $\{q\} \rightarrow \neg p$

To illustrate why we have defined unfounded sets the way we have, assume that in this example $\{r_1, r_2\}$ is the only conflict set and $r_2 \prec r_1$ (in the other logics, choosing conflict sets in this arbitrary fashion does not have the same negative effect as shown here). Since $r_2 \prec r_1$, we have $D \not\approx_{MDL} p$. However, since r_3 is strict, we have $D \not\approx_{MDL} q$. But this means that we have $head(r_2) \in \mathcal{T}_n$ for some n. If we only considered the head of r_2 to determine whether p is unfounded and ignored the fact that $r_2 \prec r_1$, then we would be able to refute p. And so the operators would yield incoherent interpretations ($p \in \mathcal{T}$ and $p \in \mathcal{U}$). To prevent this, we have defined unfounded sets to explicitly consider both the bodies of conflicting rules and the rules' locations in the precedence relation.

6.10 A Proof System

The proof system for MDL is shown below.

Definition 6.49. *τ is a defeasible proof in MDL iff τ is an argument tree for D, and for each node n of τ, one of the following obtains:*

(1) n is labeled $+\delta p$ and either:

(1.1) there is a $r \in R_s[p]$ such that $body(r)$ succeeds at n, or
(1.2) there is a $r \in R_d[p]$ such that

(1.2.1) $body(r)$ succeeds at n, and
(1.2.2) for all $c \in C[r]$, then there is a $s \in c - \{r\}$ such that either

(1.2.2.1) $\{head(s)\} \cup body(s)$ fails at n, or
(1.2.2.2) $s \prec r$.

(2) n *is labeled* $-\delta p$ *and:*

(2.1) *for all* $r \in R_s[p]$, $body(r)$ *fails at* n, *and*
(2.2) *for all* $r \in R_d[p]$, *either*

(2.2.1) $body(r)$ *fails at* n, *or*
(2.2.2) *there is a* $c \in C[r]$ *such that for each* $s \in c - \{r\}$

(2.2.2.1) $\{head(s)\} \cup body(s)$ *succeeds at* n, *and*
(2.2.2.2) $s \not\prec r$.

3. n *is labeled* $-\delta p$ *and has an ancestor* m *in* τ *labeled* $-\delta p$, *and all nodes between* n *and* m *are negative defeasible assertions.*

As is proven in Appendix E, the proof system is sound *wrt* the semantics, and for finite grounded components, it is complete.

Theorem 6.50 (Soundness)**.** *Let* D *be a defeasible theory and* $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, *. . ., the sequence of interpretations created by iterating* W_D *from* $\langle \varnothing, \varnothing \rangle$.

(1) *If* $D \mid\sim_{MDL} p$, *then there exists a finite* $\alpha \geq 0$ *and a finite grounded component* $\mathcal{X}$ *of* $\mathcal{I}_{D,\alpha}$ *such that* $p \in \mathcal{T}_{\mathcal{X}}$.
(2) *If* $D \sim|_{MDL} p$, *then there exists a finite* $\alpha \geq 0$ *and a finite grounded component* $\mathcal{X}$ *of* $\mathcal{I}_{D,\alpha}$ *such that* $p \in \mathcal{U}_{\mathcal{X}}$.

Theorem 6.51 (Completeness for finite grounded components)**.** *Let* D *be a defeasible theory and* $\mathcal{I}_{D,0}, \mathcal{I}_{D,1}$, *. . ., the sequence of interpretations created by iterating* W_D *from* $\langle \varnothing, \varnothing \rangle$. *For any* $\alpha \geq 0$, *If* $\mathcal{X}$ *is a finite grounded component of* $\mathcal{I}_{D,\alpha}$, *then*

(1) *if* $p \in \mathcal{T}_{\mathcal{X}}$, *then* $D \mid\sim_{MDL} p$, *and*
(2) *if* $p \in \mathcal{U}_{\mathcal{X}}$, *then* $D \sim|_{MDL} p$.

Appendix E also shows that for *unprioritized* theories, MDL satisfies versions of both Cut and Cautious Monotony. We do not know if it holds for prioritized theories, even if the precedence relation is transitive. Note, however that when we add a rule to a theory D, then the conflict sets of D must be modified. If we add $s : \{\} \rightarrow p$ to D, then if $r \in R_{D,du}[p]$ and $c \in C_D[r]$, r should be deleted from c. If we add $s : \{\} \Rightarrow p$ to a theory, then we form an additional conflict set $c' = (c - \{r\}) \cup \{s\}$. Rule Simplification and Elimination hold in the general case.

Theorem 6.52 (Cut and Cautious Monotony for Unprioritized Theories)**.** *Let $D = \langle R_D, C_D, \varnothing\rangle$ be a defeasible theory such that $D \mid\!\approx_{MDL} p$. Let $E = \langle R_D \cup \{r_p\}, C_E, \varnothing\rangle$ where $r_p = \{\} \rightarrow p$ or $r_p = \{\} \Rightarrow p$ and C_E is described as above. Then the following hold:*

(1) $D \mid\!\approx_{MDL} q$ iff $E \mid\!\approx_{MDL} q$.
(2) $D \approx\!\mid_{MDL} q$ iff $E \approx\!\mid_{MDL} q$.

Theorem 6.53 (Rule Simplification)**.** *Let D a defeasible theory such that $D \mid\!\approx_{MDL} p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) $D \mid\!\approx_{MDL} q$ iff $E \mid\!\approx_{MDL} q$.
(2) $D \approx\!\mid_{MDL} q$ iff $E \approx\!\mid_{MDL} q$.

Theorem 6.54 (Rule Elimination)**.** *Let $D = \langle R_D, C_D, \prec_D\rangle$ be a prioritized defeasible theory such that $D \approx\!\mid_{MDL} p$. Let t be any rule such that $p \in body(t)$, and let $E = \langle R_D - \{t\}, C_E, \prec_E\rangle$, where $C_E = C_D - \{c | t \in c\}$ and $\prec_E = \prec_D - \{a \prec b | (a \prec b) \wedge t \in \{a, b\}\}$. For all $q \in Lit_D$,*

(1) $D \mid\!\approx_{MDL} q$ iff $E \mid\!\approx_{MDL} q$.
(2) $D \approx\!\mid_{MDL} q$ iff $E \approx\!\mid_{MDL} q$.

6.11 Related Work

A few (but only a few) suggestions for incorporating priorities into the WFS appear in the literature. In this section we discuss some, paying special attention to the proposal by Brewka in [Bre96]. In this system, the consequences of the semantics with priorities are a strict superset of those under the WFS for extended program, and so the semantics is a generalization of the WFS. This is a significant virtue; MDL does similarly extend SDL. Furthermore, in Brewka's system at least, the consequences can be computed in cubic time relative to the size of the original theory. This, again, is a significant virtue. The complexity of computing consequences under MDL is not known.

Ultimately, we take a somewhat pragmatic view. Since there is a translation scheme for defeasible logic into logic programs, and since the prioritized extensions to the WFS do appear to have several virtues, these may be seen as "good" ways of handling priorities. However, the semantics

such as Brewka's are typically defined as the least fixpoint of a γ^2–like operator and in general will not be recursively enumerable (recall that the WFS is not enumerable either). Rather importantly, the definitions lend themselves most naturally to bottom-up procedures for calculating the results; it appears that no top–down or goal directed procedures have been ever been proposed. To determine whether a literal appears in the prioritized well–founded model, the entire model must first be generated. In contrast, the proof system for MDL is top–down; in an implementation in which unrestricted function systems are allowed, a top–down procedure for computing consequences may be essential.

6.11.1 BREWKA'S WFS WITH DYNAMIC PREFERENCES

Brewka begins by providing an alternative but equivalent definition of the WFS for extended logic programs. In it he defines $SAFE_X(\Pi)$, the portion of program Π that is "safe" to apply given a set of literals X. Once this is defined for unprioritized programs, he modifies the definition, using the priorities to maximize the number of rules in $SAFE_X(\Pi)$.

Definition 6.55 (Defeat of a rule)**.** *Let Π be an extended logic program, r a rule of Π, and X a set of literals from Lit_Π. r is defeated by X if $body(r)^- \cap X \neq \varnothing$.*

Definition 6.56 (Reduct)**.** *Let Π be an extended logic program and X a set of literals from Lit_Π. Π_X is the set of rules not defeated by X.*

Note that the reduct as defined above is simply the reduct as we have usually been defining it, save that default literals are not deleted from rules. Brewka redefines γ so that the default literals are ignored when computing the closure of rules. In fact, Brewka defines a further nonexplosive version of the operator, denoted γ^*. The reason for this is that the least fixpoint of $\gamma_\Pi\gamma^*_\Pi$ is a closer approximation of the consequences under the answer-set semantics than is the least fixpoint of γ^2_Π.

Definition 6.57. *Let Π be an extended logic program and X a set of literals from Lit_Π. We define the following:*

(1) $Mon(\Pi)$ *is the program obtained by deleting each default literal from* Π.

(2) $Cl(\Pi) = Mon(\Pi) \uparrow \omega$.
(3) $Cn(\Pi) = Cl(\Pi)$ *if* $Cl(\Pi)$ *consistent; otherwise* $Cn(\Pi) = Lit_{\Pi}$.
(4) $\gamma_{\Pi}(X) = Cn(\Pi_X)$.
(5) $\gamma^*_{\Pi}(X) = Cl(\Pi_X)$.
(6) $SAFE_X(\Pi) = \Pi_{Cl(\Pi_X)}$.

Informally, $SAFE_X(\Pi)$ is the set of rules not defeated by $Cl(\Pi_X)$. As the below sequence of equivalences show, given these definitions, $\gamma_{\Pi}\gamma^*_{\Pi}(X)$ can be recast as $Cl(SAFE_X(\Pi))$.

(1) $\gamma^*_{\Pi}(X) = Cl(\Pi^X) = Cl(\Pi_X)$.
(2) $SAFE_X(\Pi) = \Pi_{Cl(\Pi_X)}$.
(3) $\gamma_{\Pi}(\gamma^*_{\Pi}(X)) = \gamma_{\Pi}(Cl(\Pi_X))$.
(4) $\gamma_{\Pi}(\gamma^*_{\Pi}(X)) = Cn(\Pi_{Cl(\Pi_X)})$.
(5) $\gamma_{\Pi}(\gamma^*_{\Pi}(X)) = Cn(SAFE_X(\Pi))$.

Definition 6.58. $\Gamma^*_{\Pi}(X) = \gamma_{\Pi}(\gamma^*_{\Pi}(X)) = Cn(SAFE_X(\Pi))$.

Definition 6.59 (Well-Founded Model)**.** *Let* Π *be an unprioritized extended logic program. The well-founded model of* Π *is* $lfp(\Gamma^*_{\Pi})$.

Informally, to compute the well-founded model, we start with $X_0 = \{\}$ and then determine the set of rules that are "safe" to apply relative to X. The closure of these rules is X_1. The limit of the sequence X_0, X_1, ... is the well-founded model. To incorporate priorities, Brewka uses the priority relation (which in his semantics is defined in the program itself and is thereby dynamic) to fine tune the set of safe rules.

Definition 6.60. *Let* R *be a set of rules and* r *a rule.*

$$Dom(r, R) = \{r' |\ r' \prec r \textit{ and } Cl(r \cup R) \textit{ defeats } r'\}.$$

$Dom(r, R)$ is the set of rules that r and R dominate. If R is a set of rules known to be acceptable and r is preferred to r' and r' is defeated by $Cl(r \cup R)$, then r' is dominated by r and R. Strict rules are always acceptable.

The set of safe rules *wrt* X, here written $SAFE^{pr}_X(\Pi)$, is computed iteratively by building up sets R_0, R_1,. . ..

Definition 6.61. *Let* Π *be an extended program and* X *a set of literals.* $SAFE_X^{pr}(\Pi) = \bigcup_{i=0}^{\infty} R_i$, *where*

(1) $R_0 = \{\}$
(2) $R_n = \{r |\ r$ *is not defeated by* $Cl(\Pi_X - Dom(r, R_{n-1}))\}$.

The well-founded model is defined in the same way, using $SAFE_X^{pr}(\Pi)$ in place of $SAFE_X(\Pi)$.

6.11.2 Others Formalisms

Another attempt to prioritize the WFS for extended logic programs is found in [WZL00]. The preference relation is transitive and irreflexive and defined over strict and defeasible rules (recall that this not allowed in defeasible logic). Like Brewka's formalism, the semantics reduces to the simpler WFS when $\prec = \varnothing$. In [SW02] it is argued that the system of [WZL00] is in a sense too skeptical and an alternative system is proposed. Both of the systems of [WZL00] and [SW02] are, like Brewka's, based upon antimonotone operators, and no top–down mechanisms for computing consequences are discussed.

A survey of preference handling in NMR formalisms, which included a discussion of preferences in the WFS, is found in [DSTW04]. An implementation of computing answer sets of ordered logic programs is described in [DST01]. The system, called `plp`, translates ordered programs into consequence preserving unordered programs which can then be processed by general answer set solvers. It is stated there that the preferred answer-set semantics of [WZL00] can be computed using `plp` (though it is not explicitly stated that the well-founded model based on the same operators can be computed). The system is implemented as a front end for `dlv`. As such, variables but not function symbols are allowed. A separate discussion of compiling prioritized programs into unprioritized programs, again in the context of answer–set programming and again using `dlv`, appears in [EFLP03]. The original program is encoded as a set of facts which is then processed by a meta-interpreter. The well-founded semantics is not directly discussed.

6.11.3 DISCUSSION

The crucial point to make about the above semantics of [Bre96], [WZL00], and [SW02], is that all are explicitly based upon an alternating fixpoint procedure. At issue ultimately is what, at a given, point, constitute the "safe" rules of the program, *i.e.* the rules that we know are OK to apply. This most evident in Brewka's semantics. To determine if a rule is safe *wrt* some set X of literals, we compute the reduct of Π *wrt* X and then the closure of some selected subset of this reduct.

Observe that determining safety in this fashion amounts to a variety of "consistency check", and it is precisely this feature that leads to the failure of semi-decidability. The same problem is present in [WZL00] and [SW02]. We can't compute the closure of the first application of $\gamma_\Pi(X)$, and so we can't say if a given rule r appears in $\Pi^{\gamma_\Pi(X)}$.

In MDL, conflicts are explicitly specified in an attempt to get around this problem. In some cases, such as that shown below, this appears to be a useful maneuver. However, we do not know if in practice this will be of much computational benefit or whether it simply becomes too difficult to be of use.

Example 6.62.

(1) $\{\} \Rightarrow p(0)$
(2) $\{p(X)\} \rightarrow p(f(X))$
(3) $\{p(X)\} \rightarrow r(X)$
(4) $\{r(f(X))\} \rightarrow r(X)$
(5) $\{r(X)\} \rightarrow \neg q$
(6) $\{\} \Rightarrow q$
$1 \prec 6$

If all conflicts are ignored in the above example, then both q and $\neg q$ are derivable from the theory. The grounding of the above theory is infinite and the closure of D would not be reached until ω (In other words, the closure of $D_0^{\mathcal{T}}$ is not finite, and so we would need to wait forever to know if any defeasible rule is "safe" to use). Furthermore, there are an infinite number of finite cycle-free paths of the form

$$p(0), p(f(0)), \ldots, p(f^n(0)), r(f^n(0)), r(f^{n-1}(0)), \ldots, r(f(0)), r(0), \neg q$$

and in an SDL–proof tree for q, we would need to check them all. But it's easy to see that the source of conflict is the set $\{r_1, r_6\}$. If we specify this in the theory explicitly, the proof of $+\delta q$ consists of a single node.

CHAPTER 7

CONCLUDING REMARKS

In this dissertation, we have mostly been concerned with drawing parallels between defeasible logic and the well-founded semantics for logic programs, or else with modifying concepts from the WFS to suit defeasible logic. Using van Gelder's original presentation of the WFS [GRS91], we were able to formulate fixpoint semantics for the logics NDL and ADL. Soundness and (restricted) completeness results were provided. Alternating fixpoint methods for computing consequences were developed. Also, we showed how logic programs and defeasible theories, under their respective well-founded semantics, are often inter-definable. Furthermore, we studied the circumstances under which the logics satisfy the commonly endorsed properties Cut and Cautious Monotony. It was also found that neither ADL nor NDL allow reinstatement; the logics SDL and MDL were developed as a result of this.

Recall the statement attributed to Pat Hayes—that anyone inventing a new logic should be fined $1000—and Horty's comment that (as of 1989) work on over 70 distinct NMR formalisms was underway. Anecdotes like this are not uncommon. There are many logics and NMR formalisms in existence. Since none has been found to be entirely satisfactory, research continues. We do not claim that the logics presented here (certainly the logics *created* here) are without their limitations. However, we do find some consolation in the below passage from Makinson [Mak94], which serves as an adequate response to the sentiment expressed by Hayes.

> There is no uniquely preferable nonmonotonic relation $\mid\!\sim$ waiting to be discovered; there may not even be any uniquely preferable set of properties that such relations should satisfy. Rather, there is a variety of frameworks and relations that are, to varying

> degrees, coherently motivated, transparently generated, regularly behaved, easily computable, and connected to practical life (we should not presume that these qualities always swing together), thus yielding tools of inference that are, overall, more or less interesting.

In general we do not hope for perfection, but we are interested in developing useful tools of inference, and we would like to improve them when we see where improvements can be made.

With that in mind, we would like to end by briefly discussing two topics that have not been deeply discussed in previous chapters—namely, (1) the complexity of drawing conclusions from defeasible theories, and (2) the need for implementations of defeasible logic. These are the most reasonable and productive topics to explore in immediate further research.

Regarding implementations, we note that Brewka's prioritized WFS appeared in print in 1996. To our knowledge, however, no implementation of it has ever existed. The same can be said for the system described in [SW02], and though the prioritized answer set semantics described in [WZL00] is implemented in the `plp` system of [DST01], we do not know if `plp` can compute the well-founded models based on the same operators. Similarly, with the exception of BDL and BDLA, no software tools exist implementing any of the more recent defeasible logics discussed in this dissertation.

While it can be argued that these systems have merit even if they never appear in real-world applications, we strongly feel that it would be better if they are actually used. In order for this to happen, however, mature and stable implementations must exist and be publicly available. Implementations have the added benefits of introducing other researchers to the formalism in question and also bringing to light further issues of theoretical and practical import. We observe that Prolog came into existence several years before the theory behind it stabilized.

On the topic of complexity, some facts are known. For ADL and SDL, since the WFS of a finite propositional logic program can be computed in quadratic time *wrt* the size of the program, and since finite propositional defeasible theories with minimal conflict sets (and no defeaters or priorities) can be translated into logic programs in linear time, we may infer that the consequences

of those theories can be computed in quadratic time. For the semi-normal versions of these theories, the consequences under SDL and ADL can be computed in polynomial time (relative to the size of the original theory).

For arbitrary finite propositional theories, however, matters are different. When conflict sets are closed under strict rules, the complexity of computing the consequences under any of the logics NDL/ADL/SDL is not known. A similar lack of knowledge about MDL exists, too. It was shown in Chapter 5 that defeaters, extended conflict sets, and priorities can be compiled away (at least in the case of ADL and NDL). However, the compilation process is in general not modular and produces theories exponentially larger than the original. We don't know if a more tractable procedure is possible (it seems doubtful, though). As this is so, the question of reasoning with arbitrary theories remains an important one. Ultimately, the answer will determine whether explicitly encoding conflict sets as part of the theory—as we have done in all of the logics—is preferable to some other means of handling conflict (*e.g.*, performing a dynamic "consistency check"). Below, a few specific and related questions are posed which we would like answered.

Question 1: *If L is one of NDL, ADL, or SDL and $D = \langle R, C, \prec \rangle$ is a finite propositional defeasible theory with conflict sets closed under strict rules and $p \in Lit_D$, then what is the time complexity for deciding whether $D \mid\!\approx_L p$ or $D \approx\!\!\mid_L p$* wrt *the size of R, C, and $\prec$? What is the complexity* wrt *the size of R and $\prec$ alone?*

Question 2: *If $D = \langle R, C, \prec \rangle$ is a finite propositional defeasible theory and C is closed under strict rules, then how does the size of C relate to the number and length of rules in R?*

Question 3: *If $D = \langle R, C, \prec \rangle$ is a finite propositional defeasible theory and C is closed under strict rules, is there a $D' = \langle R, C', \prec \rangle$, where $C' \subset C$, such that $wfm(D) = wfm(D')$?*

Regarding Question 1, since conflict sets can be dynamically generated from the rules of the theory (recall that the theory is propositional), one really need not encode them explicitly. However, we do not know whether considering them as part of the input makes a difference in complexity.

Regarding Question 2, it appears that things are rather bad. Consider the following case: Let $B = \{p_0, p_1, \ldots\}$ be a finite set of ground atoms and consider the set R of strict rules

$$\{\{p_i, p_j\} \rightarrow \neg p_k | p_i, p_j, p_k \in B\} \cup \{\{p_i, p_j\} \rightarrow p_k | p_i, p_j, p_k \in B\}$$

Observe that the size of the theory is $O(|B|^3)$. If we close conflict sets under strict rules, each $\{p_i, p_j, p_k\} \subseteq B$ is a conflict set. Note that this implies that $\{p_i\}$ is itself a conflict set ($i = j = k$), as is any set $\{p_i, p_j\}$ ($k = i$ or $k = j$). But since $\{p_i, p_j, p_k\}$ is a conflict set and

$$\{p_m, p_n\} \rightarrow p_i$$

is a rule in R, we have $\{p_m, p_n, p_j, p_k\}$ as a conflict set. In general, given a conflict set c, we may take any element $p_i \in c$ and replace it with two elements p_m and p_n. As this is so, any set $c \in 2^{|B|}$ is a conflict set. However, we do not know whether the sorts of theories encountered in practice would give rise to such an intractably large set of conflict sets.

Regardless of the initial size of C, if we do explicitly encode conflict sets as part of the theory, we would like to be able to prune away as many sets as possible that do no useful work. Such is the motivation behind Question 3. The difficulty is that we do not know if there is some *easy* way to identify the useless conflict sets. In the case of finite theories we can use brute force—computing the consequences of a theory before and after deleting a conflict set—but this is not a particularly attractive option. If no process can be found for general defeasible theories, at least we would like to know if there are any useful classes of theory where the conflict sets can be quickly cut down to a workable size.

BIBLIOGRAPHY

[AB01] G. Antoniou and D. Billington. 2001. Relating defeasible and default logic. *Australian Joint Conference on Artificial Intelligence*, Adelaide, Australia, December 2001. M. Stumptner, D. Corbett, and M. J. Brooks, Eds. Springer, 13–24.

[ABG+00] G. Antoniou, D. Billington, G. Governatori, M. J. Maher, and A. Rock. 2000. A family of defeasible reasoning logics and its implementation. *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000)*, W. Horn, Ed. IOS Press, Amsterdam, 459–463.

[ABGM00b] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. 2000. A flexible framework for defeasible logics. In *Proceedings of 8th International Workshop on Non-Monotonic Reasoning*, Breckenridge, Colorado, April 2000.

[ABGM00c] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. 2000. Representation results for defeasible logic. *ACM Transactions on Computational Logic*, 2(2):255–287.

[ABGM06] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher. 2006. Embedding defeasible logic into logic programming. *Theory and Practice of Logic Programming*, 6(6):703–735.

[ABM98] G. Antoniou, D. Billington, and M. J. Maher. 1998. Normal forms for defeasible logic. In *Proceedings of the 1998 joint international conference and symposium on Logic programming*, Manchester, UK, 160–174.

[AP96] J. J. Alferes and L. M. Pereira. 1995. *Reasoning with Logic Programming*. Springer-Verlag Inc., New York, NY.

[ADP95] J. J. Alferes, C. V. Damasio, and L. M. Pereira. 1995. A logic programming system for nonmonotonic reasoning. *Journal of Automated Reasoning* , 14(1):93–147.

[AM02] G. Antoniou and M. J. Maher. 2002. Embedding defeasible logic into logic programs. *Proceedings of the 18th International Conference on Logic Programming (ICLP 2002)*, Copenhagen, Denmark, July 2002. P. J. Stuckey, ed. Springer, 393–404.

[Ant06] G. Antoniou. 2006. Defeasible reasoning: A discussion of some intuitions. *International Journal of Intelligent Systems* 21(6):545–558.

[BG94] C. Baral and M. Gelfond. 1994. Logic Programming and Knowledge Representation. *Journal of Logic Programming* , 19–20:73-148.

[Bar03] C. Baral. 2003. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, New York, NY.

[BCN90] D. Billington, K. De Coster, and D. Nute. 1990. A modular translation from defeasible nets to defeasible logics. *Journal of Experimental and Theoretical Artificial Intelligence*, 2:151–177.

[Bil93] D. Billington. Defeasible logic is stable. 1993. *Journal of Logic and Computation* , 3(4):379–400.

[BL03] Y. Babovich and V. Lifschitz. 2003. *Computing Answer Sets Using Program Completion*. unpublished draft.

[Bre96] G. Brewka. 1996. Well-founded semantics for extended logic programs with dynamic preferences. *Journal of Artificial Intelligence Research*, 4:19–36.

[Bre01] G. Brewka. 2001. On the relationship between defeasible logic and well-founded semantics. *Proceedings of the 6th International Conference on Logic Programming and Nonmonotonic Reasoning*, T. Eiter, W. Faber, and M. Truszczynski, Eds. Springer–Verlog, London, 121–132.

[BS93] C. Baral and V. S. Subrahmanian. 1993. Dualities between alternative semantics for logic programming and nonmonotonic reasoning. *Journal of Automated Reasoning*, 10(3):399–420.

[BSF95] K. A. Berman, J. S. Schlipf, and J. V. Franco. Computing well-founded semantics faster. *Logic Programming and Nonmonotonic Reasoning, Third International Conference*, Lexington, KY, June 1995, V. W. Marek and A. Nerode, Eds. Springer, 113–126.

[Cam06] M. Caminada. 2006. Well-founded semantics for semi-normal extended logic programs. *Proceedings of the 11th International Workshop on Nonmonotonic Reasoning (NMR'06)*, J. Dix and A. Hunter, Eds. Number IFI-06-04 in Technical Report Series. Clausthal University of Technology, Institute for Informatics.

[Cit06] Citeseer. 2006. Most cited source documents in the citeseer continuity database as of september 2006. Available from `http://citeseer.ist.psu.edu/source.html` [Accessed May 7, 2007].

[CNV97] M. A. Covington, D. Nute, and A. Vellino. 1997. *Prolog programming in depth*. Prentice-Hall, Upper Saddle River, NJ, 1997.

[CR93] A. Colmerauer and P. Roussel. 1993. The birth of prolog. *The second ACM SIGPLAN conference on History of programming languages*, Cambridge, MA. ACM Press, New York, NY, 37–52.

[CS01] B. Cui and T. Swift. 2001. Preference logic grammars: Fixed-point semantics and application to data standardization. *Artificial Intelligence*, 138(1-2):117147.

[Dam96] C. V. Damśio. 1996 *Paraconsistent Extended Logic Programming with Constraints*. PhD. Thesis. Faculdade de Ciencias e Tecnologia da Universidade Nova de Lisboa.

[DP97] C. V. Damásio and L.M. Pereira. 1997. A paraconsistent semantics with contradiction support detection. *Proceedings of the 4th International Conference on Logic Programming and Nonmonotonic Reasoning*, J. Dix, U. Furbach, and A. Nerode, Eds. Springer, 224–243.

[DST01] J. Delgrande, T. Schaub, and H. Tompits. 2001. plp: A Generic Compiler for Ordered Logic Programs. *Proceedings of the 6th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR-01)*, T. Eiter, W. Faber, and M. Truszczynski, Eds. Springer, 411–415.

[DSTW04] J. Delgrande, T. Schaub, H. Tompits, and K. Wang. 2004. A Classification and Survey of Preference Handling Approaches in Nonmonotonic Reasoning *Computational Intelligence*, 20(2):308–334.

[DG84] W. F. Dowling and J. H. Gallier. 1984. Linear-timeformulashms for testing the satisfiability of propositional horn formulas. *Journal of Logic Programming*, 1(3):267–284.

[Dix94] J. Dix. 1994. Semantics of Logic Programs: Their Intuitions and Formal Properties. An Overview. *Logic, Action and Information. Proceedings of the Konstanz Colloquium in Logic and Information (LogIn '92)*, A. Fuhrmann and H. Rott, Eds. DeGruyter.

[Dix94] J. Dix, L. Pereira, and T. Przymusinski. 1997. Prolegomena to Logic Programming for Non-Monotonic Reasoning. *Nonmonotonic Extensions of Logic Programming*, J. Dix and L. Pereira and T. Przymusinski, Eds. Springer, Berlin, 1–36.

[DL91] J. Dorosh and R. P. Loui. 1991 Edited transcription of the workshop on defeasible reasoning with specificity and multiple inheritance, St. Louis, April 1989. *SIGART Bulletin*, 2(1):3–51.

[Don99] S. Donnelly. 1999 *Semantics, Soundness, and Incompleteness for a Defeasible Logic*. Masters thesis, The University of Georgia.

[Dun95] P. M. Dung. 1995. An argumentation-theoretic foundation for logic programming. *Journal of Logic Programming*, 22(2):151–171.

[EFK+00] T. Eiter, W. Faber, C. Koch, N. Leone, and G. Pfeifer. 2000. Declarative problem-solving using the DLV system. *Logic-based artificial intelligence*. Kluwer Academic Publishers, Norwell, MA, 79–103.

[EFLP03] T. Eiter, W. Faber, N. Leone, and G. Pfeifer. 2003. title = Computing preferred answer sets by meta-interpretation in answer set programming. *Theory and Practice of Logic Programming*. 3(4–5): 463–498.

[End72] H. Enderton. 1972. *A Mathematical Introduction to Logic*. Academic Press, New York, NY.

[Gab85] D. Gabbay. 1985. Theoretical foundations for non-monotonic reasoning in expert systems. In *Logics and models of concurrent systems*. Springer-Verlag, New York, NY, 439–457.

[GS04] A.J. Garcia and G.R. Simari. 2004. Defeasible logic programming: an argumentative approach. *Theory and Practice of Logic Programming*, 4: 95-138.

[Gel93] A van Gelder. 1993. The alternating fixpoint of logic programs with negation. *Journal of Computer and System Sciences*, 47(1):185–221.

[GL88] M. Gelfond and V. Lifschitz. 1988. The stable model semantics for logic programming. *Proceedings of the Fifth International Conference and Symposium on Logic*

Programming, Seattle, WA, August 1988. R. A. Kowalski, K. A. Bowen, Eds. MIT Press, 1070–1080.

[GL91] M. Gelfond and V. Lifschitz. 1991 Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–386.

[GLM04] E. Giunchiglia, Y. Lierler, and M. Maratea. 2004. SAT-based answer set programming. *Proceedings of the Nineteenth National Conference on Artificial Intelligence*, . D. L. McGuinness and G. Ferguson, Eds. AAAI Press/MIT Press, 61–66.

[GMAB00] G. Governatori, M. J. Maher, G. Antoniou, and D. Billington. 2000. Argumentation semantics for defeasible logics. *Pacific Rim International Conference on Artificial Intelligence*, 27–37.

[GMAB04] G. Governatori, M. J. Maher, G. Antoniou, and D. Billington. 2004. Argumentation semantics for defeasible logic. *Journal of Logic and Computation*, 14(5):675–702.

[GRS88] A. Van Gelder, K. A. Ross, and J. Schlipf. 1991. Unfounded sets and well-founded semantics for general logic programs. In *Proceedings 7th ACM Symposium on Principles of Database Systems*, Austin, TX, 221–230.

[GRS91] A. Van Gelder, K. A. Ross, and J. Schlipf. 1991. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650.

[Hay73] P. Hayes. 1973. The frame problem and related problems in artificial intelligence. In *Artificial and Human Thinking*, A. Elithorn and D. Jones, Eds. Jossey-Bass, Inc. and Elsevier Scientific Publishing Company, 45–59.

[Her30] J.Herbrand. 1930. *Recherches sur la thorie de la dmonstration*. Ph.D. Thesis, University of Paris.

[Hog90] C. J. Hogger. 1990, *Essentials of logic programming*. Oxford University Press, New York, NY.

[Hor94] J. F. Horty. 1994. Some direct theories of nonmonotonic inheritance. In *Handbook of logic in artificial intelligence and logic programming (vol. 3): nonmonotonic reasoning and uncertain reasoning*, D. Gabbay and C. Hogger, Eds. Oxford University Press, Inc., New York, NY, 111–187.

[Hor01] J. F. Horty. 2001. Argument construction and reinstatement in logics for defeasible reasoning. *Artificial Intelligence and Law*, 9(1):1–28.

[Hor02] J. F. Horty. 2002. Skepticism and floating conclusions. *Artificial Intelligence*, 135(1-2):55–72.

[HTT87] J. F. Horty, R. H. Thomason, and D. S. Touretzky. 1990. A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence*, 42(2-3):358–363.

[Kle52] S. Kleene. 1952. *Introduction to Metamathematics*. Van Nostrand Company, Inc., New York, NY.

[Kun87] Ken Kunen. 1987. Negation in logic programming. *Journal of Logic Programming*, 4:289–308, 1987.

[Lif96] V. Lifschitz. 1996. Foundations of logic programming. In *Principles of Knowledge Representation*, G. Brewka, Ed. CSLI Publications, Stanford, CA, 69–127.

[LT00] Z. Lonc and M. Truszczynski. 2001. On the problem of computing the well-founded semantics. *Theory and Practice of Logic Programming* 2001(1): 591–609.

[LZ02] F. Lin and Y. Zhao. 2004. ASSAT: Computing answer sets of a logic program by sat solvers. *Artificial Intelligence*, 157(1–2):115–137.

[Llo87] J.W. Lloyd. 1987. *Foundations of Logic Programming*. Springer-Verlag Inc., New York, NY.

[MAB98] M. J. Maher, G. Antoniou, and D. Billington. 1998. A study of provability in defeasible logic. *AI '98: Selected papers from the 11th Australian Joint Conference on*

Artificial Intelligence on Advanced Topics in Artificial Intelligence, Brisbane, Australia, July 1998, G. Antoniou and J. K. Slaney, Eds. Springer-Verlag, London UK, 215–226.

[Mah01] M. J. Maher. 2001. Propositional defeasible logic has linear complexity. *Theory and Practice of Logic Programming*, 1(6):691–711.

[Mah02] Michael J. Maher. A model-theoretic semantics for defeasible logic. 2002. *Proceedings of Workshop on Paraconsistent Computational Logic*, H. Decker, J. Villadsen, and T. Waragai, Eds. vol. 95 of *Datalogiske Skrifter*. Roskilde University, Roskilde, Denmark, 67–80.

[Mak94] D. Makinson. 1994. General patterns in nonmonotonic reasoning. In *Handbook of logic in artificial intelligence and logic programming (vol. 3): nonmonotonic reasoning and uncertain reasoning*, D. Gabbay and C. Hogger, Eds. Oxford University Press, Inc., New York, NY, 35–110.

[McC59] J. McCarthy. 1959 Programs with common sense. *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, London, UK 1958. Her Majesty's Stationary Office, 75–91.

[McC80] J. McCarthy. 1980. Circumscription — a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1-2):27–39.

[McD05] M. McDougall. 2005. *Modeling And Analyzing Integrated Policies*. PhD Dissertation. University of Pennsylvania.

[MD80] D. V. McDermott and J. Doyle. 1980 Non-monotonic logic I. *Artificial Intelligence*, 13(1-2):41–72.

[MG99] M. J. Maher and G. Governatori. 1999. A semantic decomposition of defeasible logics. *Proceedings of the sixteenth national conference on Artificial intelligence*

and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence, Orlando FL. American Association for Artificial Intelligence, Menlo Park, CA, 299–305.

[Min74] M. Minsky. 1974. A framework for representing knowledge. Technical Report. Massachusetts Institute of Technology, Cambridge, MA.

[Min93] J. Minker. 1993. An overview of nonmonotonic reasoning and logic programming. *Journal of Logic Programming*, 17(2/3&4):95–126.

[MN06a] F. Maier and D. Nute. 2006. Ambiguity propagating defeasible logic and the well-founded semantics. *Proceedings of 10th European Conference on Logics in Artificial Intelligence*, Liverpool, UK, September 2006, M. Fisher, W. van der Hoek, B. Konev, and A. Lisitsa, Eds. Springer, 306–318.

[MN06b] F. Maier and D. Nute. 2006. Relating defeasible logic to the well-founded semantics for normal logic programs. *Proceedings of the 11th International Workshop on Nonmonotonic Reasoning (NMR'06)*, J. Dix and A. Hunter, Eds. Number IFI-06-04 in Technical Report Series. Clausthal University of Technology, Institute for Informatics.

[Moo85] R. C. Moore. 1985. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1):75–94.

[MRA+01] M. J. Maher, A. Rock, G. Antoniou, D. Billington, and T. Miller. 2001. Efficient defeasible reasoning systems. *International Journal on Artificial Intelligence Tools*, 10(4):483–501.

[MS91] D. Makinson and K. Schechta. 1991. Floating conclusions and zombie paths: two deep difficulties in the 'directly skeptical' approach to inheritance nets. *Artificial Intelligence*, 48:199–209.

[MT91] V. W. Marek and M. Truszczynski. 1991. Autoepistemic logic. *Journal of the ACM*, 38(3):588–619.

[NS96] I. Niemelä and P. Simons. 1996. Efficient implementation of the well-founded and stable model semantics. *Proceedings of the 1996 Joint International Conference and Syposium on Logic Programming* Bonn, Germany, September 1996, M. Maher, Ed. MIT Press, 289–303.

[NBC89] D. Nute, D. Billington, and K. De Coster. 1989. Defeasible logic and inheritance hierarchies with exceptions. In *Proceedings of the Tbingen Workshop on Semantic Nets and Nonmonotonic Reasoning*, Vol. I. SNS Bericht 89-48, University of Tbingen, 69–82.

[Nut86] D. Nute. 1986. LDR: A logic for defeasible resasoning. ACMC Research Report 01-0013, The University of Georgia.

[Nut87] D. Nute. 1987. Defeasible reasoning. *Proceedings of the 20th Hawaii International Conference on System Science*, IEEE Press, 470–477.

[Nut88] D. Nute. 1988. Defeasible reasoning and decision support systems. *Decision Support Systems*, 4(1):97–110.

[Nut94] D Nute. 1994. Defeasible logic. In *Handbook of logic in artificial intelligence and logic programming (vol. 3): nonmonotonic reasoning and uncertain reasoning*, D. Gabbay and C. Hogger, Eds. Oxford University Press, Inc., New York, NY, 353–395.

[Nut97] D. Nute. 1997. Apparent obligation. *Defeasible Deontic Logic*, D. Nute, Ed. Kluwer Academic Publishers, Dordrecht, Netherlands, 287–315.

[Nut01] D. Nute. 2001. Defeasible logic: Theory, implementation, and applications. *Proceedings of the 14th International Conference on Applications of Prolog (INAP 2001)*, Tokyo, Japan, 2001. IF Computer Japan, 87–114.

[PA92] L. M. Pereira and J. J. Alferes. 1992. Well founded semantics for logic programs with explicit negation. *Proceedings of the 10th European conference on Artificial intelligence*, John Wiley & Sons, New York, NY, 102–106.

[Pol74] J. L. Pollock. 1974. *Knowledge and Justification*. Princeton University Press.

[Pol87] J. L. Pollock. 1987. Defeasible reasoning. *Cognitive Science*, 11(4):481–518.

[Pra02] H. Prakken. 2002. Intuitions and the modelling of defeasible reasoning: some case studies. *Proceedings of the Ninth International Workshop on Non-Monotonic Reasoning*, Toulouse, France, April 2002, S. Benferhat and E. Giunchiglia, Eds. 91–99.

[Prz91] T. C. Przymusinski. 1991, Stable semantics for disjunctive programs. *New Generation Computing*, 9(3/4):401–424.

[PS97] H. Prakken and G. Sartor. 1997. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7(1):25–75.

[Rei77] R. Reiter. 1977. On closed world data bases. *Logic and Data Bases*, 55–77. Reprinted in *Readings in nonmonotonic reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, 300–310.

[Rei80] R. Reiter. 1980. A logic for default reasoning. *Artificial Intelligence*, 13:81–132.

[Rum02] D. Rumsfeld. 2002. Secretary Rumsfeld Press Conference at NATO Headquarters, Brussels, Belgium, June 6 2002. Available from `http://www.defenselink.mil/transcripts/transcript.aspx?transcriptid=3490` [Accessed June 20, 2007].

[SW02] T. Schaub and K. Wang. 2002. Preferred well-founded semantics for logic programming by alternating fixpoints. In S. Benferhat, E. Giunchiglia, editors, *Proceedings*

of the Ninth International Workshop on Non-Monotonic Reasoning (NMR'02), 238-246.

[Sch94] G. Schurtz. 1994. Defeasible reasoning based on constructive and cumulative rules. *Philosophy and Cognitive Sciences*, R. Casati, B. Smith, and G. White, Eds. Holder-Pichler-Tempsky, 297–310.

[Sha97] M. Shanahan. 1997. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. MIT Press.

[SSW94] K. F. Sagonas, T. Swift, and D. S. Warren. 1994. XSB as an efficient deductive database engine. *SIGMOD Conference*, Minneapolis, MN, June 1994, R. T. Snodgrass and M. Winslett, Eds. ACM Press, 442–453.

[Ste89] L. A. Stein. 1989. Skeptical inheritance: Computing the intersection of credulous extensions. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Detroit, MI, August 1989, N. S. Sridharan, Ed. Morgan Kaufmann, 1153–1160.

[Ste90] L. A. Stein. 1990. *Resolving Ambiguity in Nonmonotonic Reasoning*. PhD Dissertation. Brown University.

[Tar55] A. Tarski. 1955. A lattice theoretic fixpoint theorem and its application. *Pacific Journal of Mathematics*, 5:285–305.

[Tar02] A. Tarski. 1936 (trans. 2002). On the concept of following logically. M. Stroinska and D. Hitchcock, Trans. *History and Philosophy of Logic*, 23(3):155 – 196.

[THT87] D. S. Touretzky, R. H. Thomason, and J. F. Horty. 1987. A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, Milan, Italy, August 1987, J. McDermott, Ed. Morgan Kaufmann, 476–482.

[TTH91] D. S. Touretzky, R. H. Thomason, and J. F. Horty. 1991. A skeptic's menagerie: Conflictors, preemptors, reinstaters, and zombies in nonrnonotonic inheritance. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Sydney, Australia, August 1991, J. Mylopoulos and R. Reiter Eds. Morgan Kaufmann, 478–485.

[vEK76] M. van Emden and R. A. Kowalski. 1976. The semantics of predicate logic as a programming language. *Journal of the ACM*, 23(4):733–742.

[WZL00] K. Wang, L. Zhou and F. Lin. 2000. Alternating Fixpoint Theory for Logic Programs with Priority. *CL '00: Proceedings of the First International Conference on Computational Logic*, Springer-Verlag, London, UK, 164–178.

[Wit90] C. Witteveen. 1990. Partial semantics for truth maintenance — a compositional approach. *JELIA '90: Proceedings of the European Workshop on Logics in AI*, J. van Eijck, Ed. Springer-Verlag, London, UK, 544–561.

[ZFB96] U. Zukowski, B. Freitag, and S. Brass. 2001. Transformation-based bottom-up computation of the well-founded model. *Theory and Practice of Logic Programming*, 1(5):497–538.

APPENDIX A

EXAMPLES COMPARING NDL, ADL, BDL, AND BDLA

Below are several examples showing the consequences of theories according to several defeasible logics. Note that no cycles occur in the examples used to arrive at the results, and so the systems do not agree even on theories without cycles. NDL_{ext} and ADL_{ext} are used to indicate that conflict sets are closed in those logics; the absence of the subscripts means that conflict sets are taken as minimal. A question mark in a cell corresponding to a given logic and literal indicates that the logic can neither prove nor refute the literal. After the examples two proofs relating BDLA to NDL and ADL are given.

A.1 EXAMPLES

$r_1 : \{\} \Rightarrow p$
$r_2 : \{p\} \rightarrow q$
$r_3 : \{q\} \Rightarrow r$
$r_4 : \{\} \Rightarrow \neg q$
$r_5 : \{\} \Rightarrow \neg r$

	p	q	r	$\neg q$	$\neg r$
BDL	$+\delta\ p$	$-\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$+\delta\ \neg r$
$BDLA$	$+\delta\ p$	$-\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$-\delta\ \neg r$
NDL	$+\delta\ p$	$+\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$-\delta\ \neg r$
NDL_{ext}	$-\delta\ p$	$-\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$+\delta\ \neg r$
ADL	$+\delta\ p$	$+\delta\ q$	?	$-\delta\ \neg q$	?
ADL_{ext}	?	?	?	?	?

Table A.1: Example 1 and Results

$r_1 : \{\} \Rightarrow p$
$r_2 : \{p\} \rightarrow q$
$r_3 : \{q\} \rightarrow r$
$r_4 : \{\} \rightarrow -r$

	p	q	r	$\neg r$
BDL	$+\delta\ p$	$+\delta\ q$	$-\delta\ r$	$+\delta\ \neg r$
$BDLA$	$+\delta\ p$	$+\delta\ q$	$-\delta\ r$	$+\delta\ \neg r$
NDL	$+\delta\ p$	$+\delta\ q$	$+\delta\ r$	$+\delta\ \neg r$
NDL_{ext}	$-\delta\ p$	$-\delta\ q$	$-\delta\ r$	$+\delta\ \neg r$
ADL	$+\delta\ p$	$+\delta\ q$	$+\delta\ r$	$+\delta\ \neg r$
ADL_{ext}	$-\delta\ p$	$-\delta\ q$	$-\delta\ r$	$+\delta\ \neg r$

Table A.2: Example 2 and Results

$r_1 : \{\} \Rightarrow -q$
$r_2 : \{\} \rightarrow q$
$r_3 : \{q\}, \neg q \Rightarrow p$
$r_4 : \{\} \Rightarrow -p$

	p	q	$\neg p$	$\neg q$
BDL	$-\delta\ p$	$+\delta\ q$	$+\delta\ \neg p$	$-\delta\ \neg q$
$BDLA$	$-\delta\ p$	$+\delta\ q$	$-\delta\ \neg p$	$-\delta\ \neg q$
NDL	$-\delta\ p$	$+\delta\ q$	$+\delta\ \neg p$	$-\delta\ \neg q$
NDL_{ext}	$-\delta\ p$	$+\delta\ q$	$+\delta\ \neg p$	$-\delta\ \neg q$
ADL	$-\delta\ p$	$+\delta\ q$	$+\delta\ \neg p$	$-\delta\ \neg q$
ADL_{ext}	$-\delta\ p$	$+\delta\ q$	$+\delta\ \neg p$	$-\delta\ \neg q$

Table A.3: Example 3 and Results

$r_1 : \{\} \Rightarrow p$
$r_2 : \{\} \Rightarrow t$
$r_3 : \{p\} \rightarrow r$
$r_4 : \{p\} \Rightarrow -t$
$r_5 : \{\} \rightarrow -r$

	p	r	t	$\neg r$	$\neg t$
BDL	$+\delta\ p$	$-\delta\ r$	$-\delta\ t$	$+\delta\ \neg r$	$-\delta\ \neg t$
$BDLA$	$+\delta\ p$	$-\delta\ r$	$-\delta\ t$	$+\delta\ \neg r$	$-\delta\ \neg t$
NDL	$+\delta\ p$	$+\delta\ r$	$-\delta\ t$	$+\delta\ \neg r$	$-\delta\ \neg t$
NDL_{ext}	$-\delta\ p$	$-\delta\ r$	$+\delta\ t$	$+\delta\ \neg r$	$-\delta\ \neg t$
ADL	$+\delta\ p$	$+\delta\ r$	?	$+\delta\ \neg r$	?
ADL_{ext}	$-\delta\ p$	$-\delta\ r$	$+\delta\ t$	$+\delta\ \neg r$	$-\delta\ \neg t$

Table A.4: Example 4 and Results

$r_1 : \{\neg t\} \Rightarrow q$
$r_2 : \{\neg t\} \rightarrow q$
$r_3 : \{\} \Rightarrow -t$
$r_4 : \{\} \Rightarrow -q$
$r_5 : \{\} \Rightarrow r$
$r_6 : \{q\} \rightarrow -r$

	q	r	$\neg q$	$\neg r$	$\neg t$
BDL	$-\delta\ q$	$+\delta\ r$	$-\delta\ \neg q$	$-\delta\ \neg r$	$+\delta\ \neg t$
$BDLA$	$-\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$-\delta\ \neg r$	$+\delta\ \neg t$
NDL	$+\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$+\delta\ \neg r$	$+\delta\ \neg t$
NDL_{ext}	$-\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$-\delta\ \neg r$	$-\delta\ \neg t$
ADL	$+\delta\ q$	$-\delta\ r$	$-\delta\ \neg q$	$+\delta\ \neg r$	$+\delta\ \neg t$
ADL_{ext}	?	?	?	?	?

Table A.5: Example 5 and Results

$r_1 : \{\neg r\} \rightarrow q$
$r_2 : \{\} \rightarrow -q$
$r_3 : \{\} \Rightarrow t$
$r_4 : \{\} \Rightarrow -r$
$r_5 : \{p\} \rightarrow -t$
$r_6 : \{q\} \Rightarrow p$
$r_7 : \{\} \Rightarrow t$
$r_8 : \{\neg t\} \rightarrow p$

	p	q	t	$\neg q$	$\neg r$	$\neg t$
BDL	$-\delta\ p$	$-\delta\ q$	$+\delta\ t$	$+\delta\ \neg q$	$+\delta\ \neg r$	$-\delta\ \neg t$
$BDLA$	$-\delta\ p$	$-\delta\ q$	$-\delta\ t$	$+\delta\ \neg q$	$+\delta\ \neg r$	$-\delta\ \neg t$
NDL	$+\delta\ p$	$+\delta\ q$	$-\delta\ t$	$+\delta\ \neg q$	$+\delta\ \neg r$	$+\delta\ \neg t$
NDL_{ext}	$-\delta\ p$	$-\delta\ q$	$+\delta\ t$	$+\delta\ \neg q$	$-\delta\ \neg r$	$-\delta\ \neg t$
ADL	$+\delta\ p$	$+\delta\ q$	$-\delta\ t$	$+\delta\ \neg q$	$+\delta\ \neg r$	$+\delta\ \neg t$
ADL_{ext}	$-\delta\ p$	$-\delta\ q$	$+\delta\ t$	$+\delta\ \neg q$	$-\delta\ \neg r$	$-\delta\ \neg t$

Table A.6: Example 6 and Results

$r_1 : \{\} \Rightarrow p$ $r_2 : \{\} \Rightarrow -p$ $r_3 : \{p\} \Rightarrow q$ $r_4 : \{\} \Rightarrow -q$				
	p	q	$\neg p$	$\neg q$
BDL	$-\delta\ p$	$-\delta\ q$	$-\delta\ \neg p$	$+\delta\ \neg q$
$BDLA$	$-\delta\ p$	$-\delta\ q$	$-\delta\ \neg p$	$-\delta\ \neg q$
NDL	$-\delta\ p$	$-\delta\ q$	$-\delta\ \neg p$	$+\delta\ \neg q$
NDL_{ext}	$-\delta\ p$	$-\delta\ q$	$-\delta\ \neg p$	$+\delta\ \neg q$
ADL	?	?	?	?
ADL_{ext}	?	?	?	?

Table A.7: Example 7 and Results

A.2 PROOFS OF P7 AND P8

P7. *If* $+\delta p \in Cl_{BDLA}(D)$*, then* $+\delta p \in Cl_{ADL}(D)$.

Proof. Recall that for the sake of comparisons, the precedence relation is empty. Suppose $+\delta p \in Cl_{BDLA}$. Then there exists a $BDLA$-proof tree τ with root n labeled $+\delta p$. We induct on the length of proof P.

(Basis) The smallest tree that could be labeled $+\delta p$ has depth 2. Then either n has a child labeled $+\Delta p$ and $\{\} \rightarrow p$ is a rule D, or else n has a child labeled $-\Delta\neg p$, and $\{\} \Rightarrow p$ is a rule of D, and for all rules $s \in R[\neg p]$, $body(s)$ is not supported at n (recall that the precedence relation is empty). However, since τ only has depth 2, this implies that for each $s \in R[\neg p]$, there is a $q \in body(s)$ such that no rules for q exist in D and so a trivial ADL–refutation of q exists. We may append such trees to a node labeled $-\delta\neg p$ to form an ADL-refutation of $\neg p$. And so $+\delta p \in Cl_{ADL}(D)$.

(Induction) Suppose τ has depth k+1 and root n labeled $+\delta p$. Suppose condition 1 of Definition 3.18 (Proof in $BDLA$) obtains. Then n has a child labeled $+\Delta p$ and there exists an $r \in R_s[p]$ such that $body(r)$ strictly succeeds at n. By Proposition $P1$, $+\Delta q \in Cl_{ADL}$ for each such q. By $P5$, $+\delta q \in Cl_{ADL}$. And so there exists a strict rule r in D such that $body(r)$ is defeasibly derivable in ADL. A proof tree for $+\Delta p$ can be constructed using by attaching the proof trees for each q in $body(r)$ to a node labeled $+\Delta p$. By Proposition $P5$, $+\delta p \in Cl_{ADL}$.

Now suppose Conditions 1.1–1.3 of Definition 3.18 obtain. Then there is a child of n labeled $-\Delta\neg p$, there is a rule $r \in R^{sd}[p]$ such that $body(r)$ succeeds at n, and for all rules $s \in R[\neg p]$, there is a $a \in body(s)$ such that a child of n is labeled $-\Sigma a$, and so $-\Sigma a \in Cl_{BDLA}$. By Proposition $P4$, $-\Sigma a \in Cl_{NDL}$. By Proposition $P6$, $-\delta\neg a \in Cl_{ADL}$. Generalizing, for each rule $s \in R[\neg p]$, there exists a $a \in body(s)$ such that $-\delta\neg a \in Cl_{ADL}$. Proof trees for these exist in ADL, and so we can construct a proof for $+\delta p$ in these systems by attaching them (plus trees for rooted with $+\delta q$ for each $q \in body(r)$) to a node labeled $+\delta p$. □

P8. If $-\delta p \in Cl_{NDL}(D)$, then $-\delta p \in Cl_{BDLA}(D)$.

Proof. Let τ be a NDL-proof tree with root n labeled $-\delta p$. Then $-\delta p \in Cl_{NDL}$. The proof is by induction on the depth of τ.

(Basis) Suppose τ has depth 1. Then $R_s[p] = \varnothing$ and for all $r \in R_d[p]$, either (1) $body(r)$ fails at n or (2) there must be a $s \in R_s[\neg p]$ such that s succeeds at n. Since τ has depth 1, r cannot fail, and so (2) must hold. However, if this is so, then since τ has depth 1, $body(s) = \varnothing$. Since $R_s[p] = \varnothing$ it follows that $-\Delta p \in Cl_{BDLA}$. Thus, the tree with root $-\delta p$ and children labeled $-\Delta p$ and $+\Delta\neg p$ forms a valid tree of $BDLA$. And so $-\delta p \in Cl_{BDLA}$.

(Induction) Suppose the hypothesis holds for all trees of depth less than k and suppose τ has depth k. For all $r \in R_s[p]$, there is a $q \in body(r)$ and a child m of n labeled $-\delta q$ (recall that cycles do not occur in D and so failure-by-looping is not an issue). Each m is the head of a NDL-proof of depth less than k and so by inductive hypothesis, $-\delta q \in Cl_{BDLA}$. For all defeasible rules $r \in R_d[p]$, either there is a $q \in body(r)$ and a child m of n labeled $-\delta q$ or there exists a $s \in R[\neg p]$ such that $body(s)$ succeeds at n. If the former, then (as before) by inductive hypothesis, $-\delta q \in Cl_{BDLA}$. If the latter, then $+\Sigma a \in Cl_{NDL}$. By Proposition $P3$, $+\Sigma a \in Cl_{BDLA}$.

Since for every $r \in R_s[p]$ there is a $q \in body(r)$ such that $-\delta q \in Cl_{NDL}$, it follows that $-\Delta q \in Cl_{NDL}$. And so $-\Delta p \in Cl_{NDL}$. By $P2$, $-\Delta p \in Cl_{BDLA}$. From the above discussion, for every $r \in R_d[p]$, there is a $q \in body(r)$ such that $-\delta q \in Cl_{NDL}$, or there is a rule s such that for all $a \in body(s)$, $+\Sigma a \in Cl_{BDLA}$. From this, it can be seen that the definition for $+\delta p$ has been satisfied for $BDLA$. And so $-\delta p \in Cl_{BDLA}$. □

APPENDIX B

A DIRECT EMBEDDING OF NDL INTO LOGIC PROGRAMS

B.1 TRANSLATING A DEFEASIBLE THEORY INTO A LOGIC PROGRAM

In this appendix we present a method for translating a defeasible theory into a normal logic program. The scheme given here is a somewhat streamlined version of that given in [MN06a] and is based upon [AM02], which provides a similar translation of Billington's logic. In our opinion, the translation presented here is not as graceful as that presented Chapter 5, and we feel it shows no deep relation between defeasible logic and logic programming. It's primary virtue is that it is designed specifically for NDL (recall that the correspondence in Chapter 5 is between ADL and the WFS); the well-founded model of the program corresponds to the consequences of NDL. Furthermore, the translation incorporates priorities, defeaters and extended conflict sets. It thus provides a means of implementing NDL in logic programming systems capable of reasoning with the WFS.

Let D be a finite propositional defeasible theory $\langle R, C, \prec \rangle$. The translation of D into a normal logic Π_D is shown below.

(1) If $\{q_1, q_2, \ldots, q_n\} \rightarrow p \in R_s$, add:
$p \ :- \ q_1, q_2, \ldots, q_n.$

(2) If $r : \{q_1, q_2 \ldots q_n\} \Rightarrow p \in R_d$, add:
$p \ :- \ q_1, q_2, \ldots, q_n, ok(r).$

(3) If $\{c_1, c_2, \ldots, c_m\} = C[head(r)]$, add:
$ok(r) \ :- \ ok(r, c_1), ok(r, c_2), \ldots, ok(r, c_m).$

(4) If $c \in C[head(r)]$ and $q \in c - \{head(r)\}$, add:
$ok(r, c) \ :- \ blocked\_literal(r, q).$

(5) If $c \in C[head(r)]$, $q \in c - \{r\}$, and $\{r_1, r_2, \ldots, r_k\} = R[q]$, add:
$blocked\_literal(r, q) \ :- \ blocked(r, r_1), blocked(r, r_2), \ldots blocked(r, r_k).$

(6) If $c \in C[head(r)]$, $head(r_i) \in c - \{head(r)\}$, and $q \in body(r_i)$, add:
$blocked(r, r_i) \ :- \ not \ q.$

(7) If $c \in C[head(r)]$, $head(r_i) \in c - \{head(r)\}$, and $r_i \prec r$, add:
$blocked(r, r_i) \ :- \ sup(r, r_i).$

(8) If $s \prec r$, then add:
$sup(r, s).$

The basic idea is as follows. In (1) the consequent of a strict rule can be derived just in case its antecedent holds. Counterarguments cannot defeat the strict rule and so need not be considered. In (2) the consequent of a defeasible rule r is derivable if its antecedent succeeds and it's *ok* to apply the rule relative to all attacks. Note that for each defeasible rule in D, exactly one rule of this form is added to Π_D. Item (3) states that it's *ok* to apply a rule r when it's *ok* relative to each conflict set in which $head(r)$ participates. Note that there will always be at least one conflict set. By (4) It's *ok* to apply rule r with respect to a conflict set c if there's a blocked literal in the set $c - \{head(r)\}$ (note that a rule of this form will exist for each $q \in c - \{head(r)\}$). By (5) a rule r blocks a literal q if r blocks every rule $s \in R[q]$. According to (6), r blocks s if s has a subgoal that cannot be derived. According to (7), r blocks s if r is superior to s. Item (8) specifies the priorities as facts of Π_D.

Note that the only place negation-as-failure occurs in the translation is in Item 6. Also, for any rule $r \in R_{sd}[p]$, there is a single corresponding logic program rule with head p. We refer to this corresponding rule as $trans(r)$.

Let D be a defeasible theory and Π_D its logic program translation, and $\mathcal{I}_{\Pi_D} = \langle \mathcal{T}_{\Pi_D,WF}, \mathcal{U}_{\Pi_D,WF} \rangle$ the program's well-founded model under the simple WFS. In the following section, we show that the above translation process is correct, in that $D \ |\!\approx_{NDL} \ p$ iff $p \in \mathcal{T}_{\Pi_D,WF}$, and $D \approx\!|_{NDL} \ p$ iff $p \in \mathcal{U}_{\Pi_D,WF}$.

B.2 A PROOF OF THE TRANSLATION'S CORRECTNESS

Lemma B.1. *Let D be a defeasible theory, Π_D its logic program translation and $\mathcal{I}_{\Pi_D}$ an interpretation for Π_D. Let $c \in C[head(r)]$ for some rule $r \in R_D$. If for each $q \in c - \{head(r)\}$, there exists a $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_{\Pi_D}$ and $s \not\prec r$. Let S be the set of these s's. Then the set*

$$\{ok(r), ok(r,c)\} \cup \{blocked\_literal(r,q) | q \in c - \{head(r)\}\} \cup \{blocked(r,s) | q \in c - \{r\} \& s \in S\}$$

constitutes an unfounded set wrt Π_D *and* $\mathcal{I}_{\Pi_D}$

Proof. Suppose for each $q \in c - \{head(r)\}$, there exists a $s \in R[q]$ such that $body(s) \subseteq \mathcal{T}_{\Pi_D}$ and $s \not\prec r$. Since $s \not\prec r$, $sup(r,s) \notin \Pi_D$ and so is unfounded relative to Π_D and all interpretations.

Note that there is a single rule for $ok(r)$ in Π_D and $ok(r,c)$ appears in both its body and in S. All of the rules for $ok(r,c)$ in Π_D have a subgoal $blocked\_literal(r,q)$ such that $blocked_literal(r,q) \in S$, where $q \in c - \{head(r)\}$. For each $q \in c - \{head(r)\}$, there is only one rule for $blocked\_literal(r,q) \in \Pi_D$ and $blocked(r,s)$ appears both in its body and S. Since $body(s) \subseteq \mathcal{T}_{\Pi_D}$, for every rule of the form

$$blocked(r,s) \ :- \ not\ a.$$

we have $a \in \mathcal{T}_{\Pi_D}$ (this is due to the manner in which Π_D is defined). Rules of this form exhaust all rules with head $blocked(r,s)$. Thus, for each literal $p \in S$, each rule for p has a classical subgoal in S or else a default subgoal in $\mathcal{T}_{\Pi_D}$. S thus satisfies the definition of unfounded set *wrt* Π_D and $\mathcal{I}_{\Pi_D}$. □

Theorem B.2 (Soundness)**.** *Let D be a defeasible theory and Π_D its translation. If $D \mid\!\approx_{NDL} p$, then $\Pi_D \mid\!\approx_{WFS} p$. If $D \not\approx_{NDL} p$, then $\Pi_D \not\approx_{WFS} p$.*

Proof. The proof is by induction on the sequence of interpretations $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$ .... We will show that for all $\alpha \geq 0$, if $p \in \mathcal{T}_{D,\alpha}$ then $p \in \mathcal{T}_{\Pi_D,WF}$ and if $p \in \mathcal{U}_{D,\alpha}$ then $p \in \mathcal{U}_{\Pi_D,WF}$. Since $\mathcal{T}_{D,0} = \mathcal{T}_{\Pi_D,0}$, the claim holds for $\alpha = 0$. Suppose it holds for $\alpha < \lambda$. We proceed by cases.

(case 1): Let $p \in \mathcal{T}_{D,\lambda}$ and suppose that λ is a successor ordinal. Then either (1) there exists a rule $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$ or (2) there exists a rule $r \in R_d[p]$ such that $body(r) \subseteq$

$\mathcal{T}_{D,\lambda-1}$ and for all $c \in C[p]$ there is a $q \in c - \{p\}$ and for all $s \in R[q]$ either $body(s) \cap \mathcal{U}_{D,\lambda-1}$ or $s \prec r$. If (1) holds, then $trans(r) = r$ and by inductive hypothesis $body(r) \subseteq \mathcal{T}_{\Pi_D,WF}$. By definition of T_Π and $\mathcal{I}_{\Pi_D,WF}$ we have $p \in \mathcal{T}_{\Pi_D,WF}$.

Suppose (2) holds. $trans(r) = p \ :- \ body(r) \cup ok(r)$. Again, $body(r) \subseteq \mathcal{T}_{\Pi_D,WF}$. Let $c \in C[p]$, q the literal referred to above, and $s \in R[q]$. If $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$, then by inductive hypothesis, $body(s) \cap \mathcal{U}_{\Pi_D,WF} \neq \varnothing$; in other words, there is a $a \in body(s)$ such that $a \in \mathcal{U}_{\Pi_D,WF}$. Since $\{blocked(r,s) \ :- \ not \ a\} \in \Pi_D$, we get $blocked(r,s) \in \mathcal{T}_{\Pi_D,WF}$ by definition of T_{Π_D}. If $s \prec r$, then $sup(r,s)$ appears as a fact of Π_D and so $blocked(r,s) \in \mathcal{T}_{\Pi_D,WF}$. Thus, for each $s \in R[q]$, $blocked(r,s) \in \mathcal{T}_{\Pi_D,WF}$. From this one may obtain $blocked\_literal(r,q) \in \mathcal{T}_{\Pi_D,WF}$ and $ok(r,c) \in \mathcal{T}_{\Pi_D,WF}$. Generalizing on c, for each $c \in C[p]$, we have $ok(r,c) \in \mathcal{T}_{\Pi_D,WF}$. From this, $ok(r) \in \mathcal{T}_{\Pi_D,WF}$. Since $body(r) \cup \{ok(r)\}) \subseteq \mathcal{T}_{\Pi_D,WF}$, it follows that $p \in \mathcal{T}_{\Pi_D,WF}$.

If λ is a limit ordinal, then there is some successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{D,\kappa}$. By inductive hypothesis, $p \in \mathcal{T}_{\Pi_D,WF}$.

(case 2): Suppose $p \in \mathcal{U}_{D,\lambda}$ and λ is a successor ordinal. Let a be any literal such that $a \in \mathcal{U}_{D,\lambda}$. Note that $\mathcal{U}_{D,\lambda}$ is the greatest unfounded set *wrt* $\mathcal{U}_{D,\lambda-1}$, and also $\mathcal{U}_{D,\lambda-1} \subseteq \mathcal{U}_{D,\lambda}$. Let $r \in R_s[a]$. Then by definition of unfounded sets for NDL, $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$. Let $r \in R_d[a]$. Then either (1) $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$ or (2) there exists a $c \in C[a]$ such that for all $q \in c - \{a\}$ there is a $s \in R[q]$ with $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ and $s \not\prec r$. If (2) then by inductive hypothesis $body(s) \subseteq \mathcal{T}_{\Pi_D,WF}$. By Lemma B.1 above, it follows that $ok(r) \in \mathcal{U}_{\Pi_D,WF}$. Generalizing on r, for every logic program rule t with head a, either $(body(t) - \{ok(r)\}) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$ or $ok(t) \in \mathcal{U}_{\Pi_D,WF}$. Generalizing on a, $\mathcal{U}_{D,\lambda}$ forms an unfounded set *wrt* Π and $\mathcal{I}_{\Pi_D}$. By definition of U_{Π_D} and $\mathcal{I}_{\Pi_D,WF}$, $\mathcal{U}_{D,\lambda} \subseteq \mathcal{U}_{\Pi_D,WF}$. In particular, $p \in \mathcal{U}_{\Pi_D,WF}$.

If λ is a limit ordinal, then there is some successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{U}_{D,\kappa}$. By inductive hypothesis, $p \in \mathcal{U}_{\Pi_D,WF}$.

□

Lemma B.3. *For each successor ordinal λ, if $ok(r) \in \mathcal{T}_{\Pi_D,\lambda}$, then for each conflict set $c \in C[head(r)]$, there exists a $q \in c - \{head(r)\}$ such that for all rules $s \in R[q]$, either (1) $body(s) \cap \mathcal{U}_{\Pi_D,\lambda-1} \neq \varnothing$ or (2) $s \prec r$ is in D.*

Proof. Suppose that $ok(r) \in \mathcal{T}_{\Pi_D,\lambda}$. Then the rule

$$ok(r) : -ok(r, c_1), ok(r, c_2), ..., ok(r, c_m).$$

appears in Π_D and each $ok(r, c_i) \in \mathcal{T}_{\Pi_D,\lambda-1}$. By monotonicity of $\mathcal{T}$, $ok(r, c_i) \in \mathcal{T}_{\Pi_D,\lambda}$. If $ok(r, c_i) \in \mathcal{T}_{\Pi_D,\lambda}$, then there must be some rule of the form

$$ok(r, c_i) : -blocked\_literal(r, q).$$

in Π_D such that $blocked\_literal(r, q) \in \mathcal{T}_{\Pi_D,\lambda-1}$, and so $blocked\_literal(r, q) \in \mathcal{T}_{\Pi_D,\lambda}$. $blocked\_literal(r, q)$ appears as the head of exactly one rule of Π_D:

$$blocked\_literal(r, q) : -blocked(r, s_1), \ldots, blocked(r, s_n).$$

It follows that each subgoal $blocked(r, s_i) \in \mathcal{T}_{\Pi_D,\lambda}$. This can occur for two reasons. (1) $\{blocked(r, s_i) :- not\ v\} \in \Pi_D$ and $v \in \mathcal{U}_{\Pi_D,\lambda-1}$ for some $v \in body(s)$, or (2) $\{blocked(r, s_i) : - sup(r, s_i)\} \in \Pi_D$ and $sup(r, s_i) \in \Pi_D$ (*i.e.*, $sup(r, s_i)$ is a fact of D). If the latter, then $s_i \prec r$. Generalizing on s_i, for each rule $s \in R_D[q]$, $body(s) \cap \mathcal{U}_{\Pi_D,\lambda-1} \neq \varnothing$ or $s \prec r$. Generalizing on c_i, for all $c \in C[head(r)]$, there exists a $q \in c-\{r\}$ such that for all $s \in R_D[q]$, $body(s) \cap \mathcal{U}_{\Pi_D,\lambda-1} \neq \varnothing$ or $s \prec r$. □

Lemma B.4. *For each successor ordinal λ, if $ok(r) \in \mathcal{U}_{\Pi_D,\lambda}$, then there exists a conflict set $c \in C[head(r)]$ in D, and for each $q \in c - \{head(r)\}$ there exists a rule $s \in R[q]$ such that $s \not\prec r$ and $body(s) \subseteq \mathcal{T}_{\Pi_D,\lambda-1}$.*

Proof. Suppose $ok(r) \in \mathcal{U}_{\Pi_D,\lambda}$. Then there is some $ok(r, c_i)$ in the rule

$$ok(r) : -ok(r, c_1), ok(r, c_2), ..., ok(r, c_m).$$

such that $ok(r, c_i) \in \mathcal{U}_{\Pi_D,\lambda}$. This implies that for every rule of the form

$$ok(r, c_i) : -blocked\_literal(r, q).$$

$blocked\_literal(r, q) \in \mathcal{U}_{\Pi_D,\lambda}$. It follows that the rule

$$blocked\_literal(r, q) \ :- \ blocked(r, s_1), blocked(r, s_2), \ldots blocked(r, s_m).$$

has some subgoal $blocked(r, s_u) \in \mathcal{U}_{\Pi_D,\lambda}$. Recall that all rules t with head $blocked(r, s_u)$ have either $not\ v$ or $sup(r, s_u)$ as their only subgoal, where $v \in body(s_u)$. Applying the definition of unfoundedness, for each rule with body $not\ v$, we have $v \in \mathcal{T}_{\Pi_D,\lambda-1}$. Since there is a rule such as t for each $v \in body(s)$, it follows that $body(s) \subseteq \mathcal{T}_{\Pi_D,\lambda-1}$. If $sup(r, s_u)$ appears as the body of some rule, then (since $blocked(r, s_u) \in \mathcal{U}_{\Pi_D,\lambda}$) it must be $sup(r, s_u) \in \mathcal{U}_{\Pi_D,\lambda}$; if that is the case, then $s \not\prec r$. Generalizing on q, every literal of $c_i - \{head(r)\}$ has a rule s with $s \not\prec r$ and $body(s) \subseteq \mathcal{T}_{\Pi_D,\lambda-1}$. □

Theorem B.5 (Completeness)**.** *Let D be a defeasible theory and Π_D its translation. If $\Pi_D \models\!\!\approx_{WFS} p$, then $D \models\!\!\approx_{NDL} p$. If $\Pi_D \not\approx_{WFS} p$, then $D \not\approx_{NDL} p$.*

Proof. The proof is by induction on the sequence of interpretations $\mathcal{I}_{\Pi_D,0}, \mathcal{I}_{\Pi_D,1} \ldots$. We will show that for all $\alpha \geq 0$, if $p \in \mathcal{T}_{\Pi_D,\alpha}$ then $p \in \mathcal{T}_{D,WF}$ and if $p \in \mathcal{U}_{\Pi_D,\alpha}$ then $p \in \mathcal{U}_{D,WF}$. Since $\mathcal{T}_{\Pi_D,0} = \mathcal{T}_{D,0}$, the claim holds for $\alpha = 0$. Suppose it holds for $\alpha < \lambda$. We proceed by cases.

(Case 1): Suppose $p \in \mathcal{T}_{\Pi_D,\lambda}$ and λ is a successor ordinal. Then there exists a rule $r \in \Pi_D$ such that $body(r) \subseteq \mathcal{T}_{\Pi_D,\lambda-1}$. If r is strict, then r appears in D unchanged and by inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$. By definition of T_D and $\mathcal{I}_{D,WF}$, $p \in \mathcal{T}_{D,WF}$.

If r is defeasible, then r has the form

$$p \ :- \ q_1, q_2, \ldots, q_m, ok(r).$$

As before, $body(r) \subseteq \mathcal{T}_{D,WF}$. In this case we also have $ok(r) \in \mathcal{T}_{\Pi_D,\lambda-1}$, and so by monotonicity $ok(r) \in \mathcal{T}_{\Pi_D,\lambda}$. By Lemma B.3, for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for each rule $s \in R[q]$, either $body(s) \cap \mathcal{U}_{\Pi_D,\lambda-1} \neq \varnothing$ or $s \prec r$. If the former, then by hypothesis,

$body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. And so we have $r \in R_d[p]$ and $body(r) \subseteq \mathcal{T}_{D,WF}$, and for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that for each rule $s \in R[q]$, either $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $s \prec r$. By definition of T_D and $\mathcal{I}_{D,WF}$, $p \in \mathcal{T}_{D,WF}$.

If λ is a limit ordinal, then there is some successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{\Pi_D,\kappa}$. By inductive hypothesis, $p \in \mathcal{T}_{D,WF}$.

(Case 2): Suppose $p \in \mathcal{U}_{\Pi_D,\lambda}$ and λ is a successor ordinal. Let a be any literal such that $a \in \mathcal{U}_{\Pi_D,\lambda}$ and let $r \in R_D[a]$, and let $t = trans(r)$. Since $a \in \mathcal{U}_{\Pi_D,\lambda}$, t has a subgoal $b \in \mathcal{U}_{\Pi_D,\lambda}$. If $b \neq ok(r)$), then $body(r) \cap \mathcal{U}_{\Pi_D,\lambda} \neq \varnothing$. If $b = ok(r)$, then by Lemma B.4, there is a $c \in C[a]$ such that for each $q \in c - \{a\}$, for some rule $s \in R[q]$, $body(s) \subseteq \mathcal{T}_{\Pi_D,\lambda-1}$ and $s \not\prec r$. By inductive hypothesis, $body(s) \subseteq \mathcal{T}_{D,WF}$. Generalizing on a, it can be seen that $\mathcal{U}_{\Pi_D,\lambda}$ is unfounded *wrt* D and $\mathcal{I}_{D,WF}$. By definition of U_D and $\mathcal{I}_{D,WF}$, $\mathcal{U}_{\Pi_D,\lambda} \subseteq \mathcal{U}_{D,WF}$. As such, $p \in \mathcal{U}_{D,WF}$.

If λ is a limit ordinal, then there is some successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_{\Pi_D,\kappa}$. By inductive hypothesis, $p \in \mathcal{T}_{D,WF}$. □

APPENDIX C

THEORY SIMPLIFICATION IN ADL AND NDL

The below proofs show that for NDL and ADL, if we delete from the body of a rule a literal known to be well-founded, then the consequences are not altered. This implies that if a rule fires, then we may delete its body without changing the consequences of a theory. It is also shown that if the body of a rule contains an unfounded literal, then we may delete the entire rule without changing the consequences of the theory. These results are of practical import, since it shows that we can use acquired information about the theory to simplify it.

A later section shows that Cut holds in general for ADL, provided the rule added to the theory is a presumption.

C.1 RULE SIMPLIFICATION

Theorem 4.46 (Rule Simplification)**.** *Let L be one of NDL or ADL, D a defeasible theory such that $D \not\approx_L p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) $D \not\approx_L q$ iff $E \not\approx_L q$.
(2) $D \not\approx\!|_L q$ iff $E \not\approx\!|_L q$.

Proof. The theorem follows directly from the below two lemmas; each proves one direction. □

Lemma C.1. *Let L be one of NDL or ADL, D a defeasible theory such that $D \not\approx_L p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) If $D \not\approx_L q$ then $E \not\approx_L q$.

(2) If $D \approx\!\!|_L\ q$ then $E \approx\!\!|_L\ q$.

Proof. The proof is by induction on the sequence $(\mathcal{I}_D)$. Suppose for all $\alpha < \lambda$, if $q \in \mathcal{T}_{D,\alpha}$ then $q \in \mathcal{T}_{E,WF}$, and if $q \in \mathcal{U}_{D,\alpha}$ then $p \in \mathcal{U}_{E,WF}$. The hypothesis is trivially satisfied for $\mathcal{T}_{D,0}$. We will treat the cases where $q \in \mathcal{T}_{D,\lambda}$ and $q \in \mathcal{U}_{D,\lambda}$ separately.

(Case 1) Suppose $q \in \mathcal{T}_{D,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $r \in R_D[q]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{E,WF}$. If r is strict, then clearly $q \in \mathcal{T}_{E,WF}$ by definition of T_E and $\mathcal{T}_{E,WF}$. Suppose r is defeasible. Let $c \in C[q]$. Then since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{q\}$ such that for all $s \in R_D[w]$ either $s \prec r$ or $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. Note that either $w = p$ or $w \neq p$.

Suppose $w = p$. Then $R_E[p] = R_D[p] - \{t\} \cup \{t'\}$. Since $D \not\approx\!\!|_L\ p$, if $body(t) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$, then there exists a $v \neq p$ such that $v \in body(t)$ and $v \in \mathcal{U}_{D,\lambda-1}$. Since $body(t') = body(t) - \{p\}$, if $body(t) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$ then $body(t') \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$, and if so then by inductive hypothesis $body(t') \cap \mathcal{U}_{E,WF} \neq \varnothing$. Similarly, for all $s \in R_D[p]$ such that $s \neq t$, either $s \prec r$ or $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$. Since $R_E[p] = R_D[p] - \{t\} \cup \{t'\}$, we have considered each rule of $R_E[p]$.

Now suppose $w \neq p$. Then $R_E[w] = R_D[w]$. As above, for all rules $s \in R_E[w]$, either $s \prec r$ or else $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Generalizing on c, we have $body(r) \subseteq \mathcal{T}_{E,WF}$ and for each conflict set $c \in C_E[q]$, there exists a $w \in c - \{q\}$ such that for each $s \in R_E[w]$, $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or else $s \prec r$. If $r = t$, then $t' \in R_E$, $body(t') \subseteq \mathcal{T}_{E,WF}$, and if $s \prec t$, then $s \prec t'$. If $r \neq t$, then $r \in R_E$. Either way, by definition of and T_E and $\mathcal{T}_{E,WF}$, $q \in \mathcal{T}_{E,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{D,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{T}_{E,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{D,\lambda}$, λ is a successor ordinal, and let a be any literal in $\mathcal{U}_{D,\lambda}$. Since $\mathcal{I}_\lambda$ is coherent and $D \not\approx\!\!|_L\ p$, $a \neq p$ and so $R_D[a] = R_E[a]$. We will consider each rule of a.

Suppose $r \in R_s[a]$. Since $a \in \mathcal{U}_{D,\lambda}$, then $body(r) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$. Since U_D is monotonic, it follows that $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$.

Suppose $r \in R_d[a]$. Then $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a $c \in C[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R[w]$ such that $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict). By the inductive hypothesis, $body(s) \subseteq \mathcal{T}_{E,WF}$. It's possible that either $s = t$ or $s \neq t$. If $s = t$, then $t' \in R_E[w]$ and $t' \not\prec r$ (for ADL, $r \prec t'$ or t' is strict), and by inductive hypothesis $body(t') \subseteq \mathcal{T}_{E,WF}$. If $s \neq t$, then $s \in R_E[w]$ and by the inductive hypothesis, $body(s) \subseteq \mathcal{T}_{E,WF}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict).

Generalizing on r, for each $r \in R_{E,s}[a]$, $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$. For each $r \in R_{E,d}[a]$, either $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a conflict set $c \in C[a]$ such that for each $w \in c - \{q\}$ there is a rule $s \in R_E[w]$ such that $body(s) \subseteq \mathcal{T}_{E,WF}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict). Generalizing on a, we see that $\mathcal{U}_{D,\lambda}$ is unfounded *wrt* E and $\mathcal{I}_{E,WF}$. It follows by definition of U_E and $\mathcal{I}_{E,WF}$ that $q \in \mathcal{U}_{E,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{D,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{U}_{E,WF}$. □

Lemma C.2. *Let L be one of NDL or ADL, D a defeasible theory such that $D \not\approx_L p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) If $E \not\approx_L q$ then $D \not\approx_L q$.
(2) If $E \approx\!\!\!\!\!\!\mid_L q$ then $D \approx\!\!\!\!\!\!\mid_L q$.

Proof. We show for all $\alpha \geq 0$, if $q \in \mathcal{T}_{E,\alpha}$, then $q \in \mathcal{T}_D$. If $q \in \mathcal{U}_{E,\alpha}$, then $q \in \mathcal{U}_{D,WF}$. The hypothesis is trivially satisfied for $\alpha = 0$. Suppose it holds for all $\alpha < \lambda$. We'll treat the case where $q \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{U}_{E,\lambda}$ separately.

(Case 1) Suppose $q \in \mathcal{T}_{E,\lambda}$ and λ is a successor ordinal. If $q = p$, then we already have $p \in \mathcal{T}_{D,WF}$ by assumption. Suppose $q \neq p$. Then there exists a rule $r \in R_E[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. Since $q \neq p$, $r \in R_D$ and by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$. If r is strict, then clearly $q \in \mathcal{T}_{D,WF}$ by definition of T_D and $\mathcal{T}_{D,WF}$.

Suppose r is defeasible. Let $c \in C[q]$. Then since $q \in \mathcal{T}_{E,\lambda}$ there exists a $w \in c - \{q\}$ such that for all $s \in R_E[w]$ either $s \prec r$ or $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. Note that every rule of $R_E[w]$

is a rule of $R_D[w]$ except possibly t'. Suppose $s = t'$. Observe that $body(t') \subset body(t)$, and so if $body(t') \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$, then $body(t) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. For every other rule $s \in R_E[w]$, it holds that $s \in R_D[w]$ and either $s \prec r$ or $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. By the inductive hypothesis, if $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$, then $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. All rules in $R_D[w]$ have thereby been accounted for. Generalizing on c, we have $body(r) \subseteq \mathcal{T}_{D,WF}$ and for each $c \in C_D[q]$ there exists a $w \in c - \{q\}$ such that for all $s \in R_D[w]$, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $s \prec r$. By definition of T_D and $\mathcal{I}_{D,WF}$, $q \in \mathcal{T}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{E,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{E,\lambda}$, λ is a successor ordinal, and let a be any literal such that $a \in \mathcal{U}_{E,\lambda}$. Since $D \not\approx_L p$, by Lemma C.1 we have $E \not\approx_L p$. By coherence, we infer that $a \neq p$, and so $R_E[a] = R_D[a]$. If $r \in R_{E,s}[a]$, then $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. Since U_E is monotonic, it follows that $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$.

Suppose $r \in R_{E,d}[a]$. Then $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a $c \in C[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R[w]$ such that $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s \not\prec r$ (for ADL, s is strict or $r \prec s$). Suppose $s = t'$. Since $body(t') \subseteq \mathcal{T}_{E,\lambda-1}$, then by inductive hypothesis, $body(t') \subseteq \mathcal{T}_{D,WF}$. By assumption, $p \in \mathcal{T}_{D,WF}$, and so $body(t) \subseteq \mathcal{T}_{D,WF}$. Since $t' \not\prec r$ (for ADL, t' is strict or else $r \prec t'$), it follows that $t \not\prec r$ (for ADL, t is strict or $r \prec t$). If $s \neq t'$ then $s \in R_D[w]$ and $s \not\prec r$ (for ADL, s is strict or $r \prec s$), and by the inductive hypothesis, $body(s) \subseteq \mathcal{T}_{D,WF}$.

Generalizing on r, for each $r \in R_{D,sd}[a]$, either $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else r is defeasible and there exists a conflict set $c \in C_D[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R_D[w]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$ (for ADL, s is strict or $r \prec s$). Generalizing on a, for all $a \in \mathcal{U}_{E,\lambda}$, the above holds. And so $\mathcal{U}_{E,\lambda}$ is unfounded *wrt* D and $\mathcal{I}_{D,WF}$. Thus $q \in \mathcal{U}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{E,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{U}_{D,WF}$. □

C.2 RULE ELIMINATION

Theorem 4.47 (Rule Elimination)**.** *Let L be one of NDL or ADL, $D = \langle R_D, C_D, \prec_D \rangle$ a prioritized defeasible theory such that $D \not\approx\!|_L\ p$. Let t be any rule such that $p \in body(t)$, and let $E = \langle R_D - \{t\}, C_D, \prec_D \rangle$. For all $q \in Lit_D$,*

(1) $D \not\approx\!|\!\approx_L q$ iff $E \not\approx\!|\!\approx_L q$.
(2) $D \not\approx\!|_L q$ iff $E \not\approx\!|_L q$.

Proof. Follows directly from the below two lemmas. □

Lemma C.3. *Let L be one of NDL or ADL, $D = \langle R_D, C_D, \prec_D \rangle$ a prioritized defeasible theory such that $D \not\approx\!|_L\ p$. Let t be any rule such that $p \in body(t)$, and let $E = \langle R_D - \{t\}, C_D, \prec_D \rangle$. For all $q \in Lit_D$,*

(1) If $D \not\approx\!|\!\approx_L q$ *then* $E \not\approx\!|\!\approx_L q$.
(2) If $D \not\approx\!|_L q$ *then* $E \not\approx\!|_L q$.

Proof. The proof is by induction on the sequence $(\mathcal{I}_D)$. Suppose for all $\alpha < \lambda$, if $q \in \mathcal{T}_{D,\alpha}$ then $q \in \mathcal{T}_{E,WF}$, and if $q \in \mathcal{U}_{D,\alpha}$ then $p \in \mathcal{U}_{E,WF}$. The hypothesis is trivially satisfied for $\mathcal{T}_{D,0}$. We will treat the cases where $q \in \mathcal{T}_{D,\lambda}$ and $q \in \mathcal{U}_{D,\lambda}$ separately.

(Case 1) Suppose $q \in \mathcal{T}_{D,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $r \in R_D[q]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{E,WF}$. Since $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$, $r \neq t$, and so $r \in R_E$. If r is strict, then clearly $q \in \mathcal{T}_{E,WF}$ by definition of T_E and $\mathcal{T}_{E,WF}$. Suppose r is defeasible. Let $c \in C[q]$. Then since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{q\}$ such that for all $s \in R[w]$ either $s \prec r$ or $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. If $t \in R_D[w]$, then $R_E[w] \subset R_D[w]$ and for each $s \in R_E[w]$, either $s \prec r$ or $body(s) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. By the inductive hypothesis, either $s \prec r$ or $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$. If $t \notin R_D[w]$, then $R_E[w] = R_D[w]$ and as before we have for all $s \in R_E[w]$, either $s \prec r$ or $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Generalizing on c, we have $body(r) \subseteq \mathcal{T}_{E,WF}$ and for each conflict set $c \in C_E[q]$, there exists a $w \in c - \{q\}$ such that for each $s \in R_E[w]$, $body(s) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or else $s \prec r$. By definition of and T_E and $\mathcal{T}_{E,WF}$, $q \in \mathcal{T}_{E,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{D,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{T}_{E,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{D,\lambda}$, λ is a successor ordinal, and let a be any literal in $\mathcal{U}_{D,\lambda}$. Suppose $r \in R_{D,s}[a]$, Since $a \in \mathcal{U}_{D,\lambda}$, then $body(r) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$. Since U_D is monotonic, it follows that $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$.

Suppose $r \in R_{D,d}[a]$, then $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a $c \in C[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R_D[w]$ such that $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict). For any such s, since $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$, it cannot be the case that $s = t$. Thus for each $w \in c - \{a\}$ there is a rule $s \in R_E[w]$ such that $body(s) \subseteq \mathcal{T}_{D,\lambda-1}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict).

Note that $R_E[a] \subseteq R_D[a]$. Generalizing on r, for each $r \in R_{E,s}[a]$, $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$. For each $r \in R_{E,d}[a]$, either $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a conflict set $c \in C[a]$ such that for each $w \in c - \{q\}$ there is a rule $s \in R_E[w]$ such that $body(s) \subseteq \mathcal{T}_{E,WF}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict). Generalizing on a,we see that $\mathcal{U}_{D,\lambda}$ is unfounded *wrt* E and $\mathcal{I}_{E,WF}$. It follows by definition of U_E and $\mathcal{I}_{E,WF}$, that $q \in \mathcal{U}_{E,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{D,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{U}_{E,WF}$. □

Lemma C.4. *Let L be one of NDL or ADL, $D = \langle R_D, C_D, \prec_D \rangle$ a prioritized defeasible theory such that $D \approx\!|_L\ p$. Let t be any rule such that $p \in body(t)$, and let $E = \langle R_D - \{t\}, C_D, \prec_D \rangle$. For all $q \in Lit_D$,*

(1) If $E \not\approx_L q$ then $D \not\approx_L q$.
(2) If $E \approx\!|_L\ q$ then $D \approx\!|_L\ q$.

Proof. The proof is by induction on the sequence $(\mathcal{I}_E)$. Suppose for all $\alpha < \lambda$, if $q \in \mathcal{T}_{E,\alpha}$ then $q \in \mathcal{T}_{D,WF}$, and if $q \in \mathcal{U}_{E,\alpha}$ then $p \in \mathcal{U}_{D,WF}$. The hypothesis is trivially satisfied for $\mathcal{T}_{E,0}$. We will treat the cases where $q \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{U}_{E,\lambda}$ separately.

(Case 1) Suppose $q \in \mathcal{T}_{E,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $r \in R_E[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D,WF}$. Since $r \in R_E$ and

$R_E \subseteq R_D$ it must be that $r \neq t$ and $r \in R_D$. If r is strict, then clearly $q \in \mathcal{T}_D$ by definition of T_D and $\mathcal{T}_{D,WF}$.

Suppose r is defeasible. Let $c \in C[q]$. Then since $q \in \mathcal{T}_{E,\lambda}$ there exists a $w \in c - \{q\}$ such that for all $s \in R_E[w]$ either $s \prec r$ or $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. By the inductive hypothesis, either $s \prec r$ or $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$. If $R_E[w] \neq R_D[w]$, then $R_E[w] = R_D[w] \cup \{t\}$, and by assumption $body(t) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Thus for all $s \in R_D[w]$, either $s \prec r$ or $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$.

Generalizing on c, we have $body(r) \subseteq \mathcal{T}_{D,WF}$ and for each conflict set $c \in C_D[q]$, there exists a $w \in c - \{q\}$ such that for each $s \in R_D[w]$, $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or else $s \prec r$. By definition of and T_D and $\mathcal{T}_{D,WF}$, $q \in \mathcal{T}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{E,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{E,\lambda}$, λ is a successor ordinal, and let a be any literal in $\mathcal{U}_{E,\lambda}$. Suppose $r \in R_{E,s}[a]$. Since $a \in \mathcal{U}_{E,\lambda}$, then $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. Since U_E is monotonic, it follows that $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$.

Suppose $r \in R_{E,d}[a]$, then $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a $c \in C[a]$ such that for each $w \in c - \{a\}$ there is a rule $s \in R_E[w]$ such that $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict). By the inductive hypothesis, for each such s, $body(s) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict).

Note that if $t \in R_D[a]$, then $R_E[a] = R_D[a] - \{t\}$, and it is already known that $body(t) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Generalizing on r, we may then infer that for each $r \in R_{D,s}[a]$, $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$. For each $r \in R_{D,d}[a]$, either $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a conflict set $c \in C[a]$ such that for each $w \in c - \{q\}$ there is a rule $s \in R_D[w]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$ (for ADL, $r \prec s$ or s is strict). Generalizing on a,we see that $\mathcal{U}_{E,\lambda}$ is unfounded *wrt* D and $\mathcal{I}_{D,WF}$. It follows by definition of U_D and $\mathcal{I}_{D,WF}$, that $q \in \mathcal{U}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{E,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{U}_{D,WF}$. □

C.3 PRESUMPTION CUT IN ADL

Regarding what happens under ADL if we add a derivable literal p as a presumption and not a fact, it seems reasonable to suspect that Cut holds. The introduction of the defeasible rule will not defeat any other rule since the new presumption is not included in the precedence relation-In the worst case, it will merely increase the number of ambiguous literals. If something can be derived/refuted in the face of this ambiguity, then its also derivable/refutable without it. A proof of this is shown below.

Theorem C.5 (Presumption Cut in ADL)**.** *Let $D = \langle R_D, C_D, \prec_D \rangle$ be a prioritized defeasible theory such that $D \mid\!\approx_{ADL} p$. Let $t = \{\} \Rightarrow p$ and $E = \langle R_D \cup \{t\}, C_D, \prec_D \rangle$. For all literals q, if $E \mid\!\approx_{ADL} q$, then $D \mid\!\approx_{ADL} q$. If $E \not\approx_{ADL} q$, then $D \not\approx_{ADL} q$.*

Proof. The proof is again by induction on the sequence $\mathcal{I}_{E,0}, \mathcal{I}_{E,0}, \ldots$. Suppose for all $\alpha < \lambda$, if $q \in \mathcal{T}_{E,\alpha}$, then $q \in \mathcal{T}_{D,WF}$, and if $q \in \mathcal{U}_{E,\alpha}$, then $q \in \mathcal{U}_{D,WF}$.

(Case 1) Let $q \in \mathcal{T}_{E,\lambda}$ and suppose that λ is a successor ordinal. Since $D \mid\!\approx_{ADL} p$, we may assume that $q \neq p$. As such $R_D[q] = R_E[q]$. Since $q \in \mathcal{T}_{E,\lambda}$, there is a rule $r \in R_{E,sd}[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$ and either r is strict or else r defeasible and for all conflict sets $c \in C[q]$ there exists a $v \in c - \{q\}$ such that for every rule $s \in R_E[v]$, either $body(s) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or else $s \prec r$.

Suppose $r \in R_{E,s}[q]$. Then $r \in R_D$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D,WF}$. By definition of T_D and $\mathcal{I}_{D,WF}$, we have $q \in \mathcal{T}_{D,WF}$.

Suppose $r \in R_{E,d}[q]$. Note that $body(t) = \varnothing$ and for any rule r' it is the case that $t \not\prec r'$ and $r' \not\prec t$. Thus it is impossible for $body(t) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$ or $t \prec r$, and so v cannot be p. Thus $R_D[v] = R_E[v]$, and so by the inductive hypothesis for every rule $s \in R_D[v]$, either $body(s) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or else $s \prec r$. Since $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$, by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$. Thus by definition of T_D and $\mathcal{I}_{D,WF}$, we have $q \in \mathcal{T}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{T}_{E,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{T}_{D,WF}$.

(Case 2) Suppose $q \in \mathcal{U}_{E,\lambda}$ and λ is a successor ordinal. Let a be any literal such that $a \in \mathcal{U}_{E,\lambda}$. We will consider each rule for a.

Let $r \in R_{E,sd}[a]$. Then either $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$, and so by the inductive hypothesis $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$, or else $r \in R_{E,d}[a]$ and there exists a conflict set $c \in C_E[a]$ such that for each $w \in c - \{a\}$ there exists a $s \in R[w]$ such that $body(s) \subseteq \mathcal{T}_{E,\lambda-1}$ and s is strict or $r \prec s$. If the latter, then by the inductive hypothesis for each $w \in c - \{a\}$ there exists a $s \in R[w]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and s is strict or $r \prec s$. Since t is not strict, and for no r' do we have $r' \prec t$, it follows that $s \neq t$. And so it must be that $s \in R_D$.

Generalizing on a, for each $a \in \mathcal{U}_{E,\lambda}$ and for each $r \in R_{E,s}[a]$, we have $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$. For each $r \in R_{E,d}[a]$, either $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{D,WF}) \neq \varnothing$ or there exists a conflict set $c \in C_E[a]$ such that for each $w \in c - \{a\}$ there exists a $s \in R_D[w]$ such that $body(s) \subseteq \mathcal{T}_{D,WF}$ and s is strict or $r \prec s$. Note that $R_D \subset R_E$, and so the above holds for any $r \in R_{D,sd}$. As this is so, it follows that $\mathcal{U}_{E,\lambda}$ is unfounded with regard to D and $\mathcal{I}_{D,WF}$. Since $q \in \mathcal{U}_{E,\lambda}$, it follows that $q \in \mathcal{U}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\kappa < \lambda$ such that $q \in \mathcal{U}_{E,\kappa}$ and so by the inductive hypothesis $q \in \mathcal{U}_{D,WF}$. □

APPENDIX D

PROOFS OF THEOREMS FOR SDL

D.1 MONOTONICITY, COHERENCE

We now prove some of the propositions claimed in Chapter 6. In the following, the operators T_D, U_D, and W_D are assumed to be defined according to SDL. When we speak of unfounded and well-founded sets of literals, for instance, it is implied that these are unfounded and well-founded under the semantics of SDL.

Lemma 6.4 (Unfounded sets are closed under union)**.** *If S_0, S_2, ... are SDL-unfounded sets* wrt *to defeasible theory D and interpretation $\mathcal{I}$, then $\bigcup_{i=0}^{\infty} S_i$ is SDL-unfounded* wrt *to D and $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$.*

Proof. Let $V = \bigcup_{i=0}^{\infty} S_i$ and $p \in V$. Suppose the $p \in S_k$ for some k. If $r \in R_s[p]$, then since S_k is unfounded, $body(r) \cap (S_k \cup \mathcal{U}) \neq \varnothing$. But since $S_k \subseteq V$, $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$. Similarly, if $r \in R_d[p]$, then either $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$ or there is a $c \in C[p]$ such that $c - \{p\} \subseteq \mathcal{T}$. Generalizing on r and p, for all $q \in V$, if $r \in R_s[q]$, $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$. If $r \in R_d[q]$, then either $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$ or there is a $c \in C[p]$ such that $c - \{q\} \subseteq \mathcal{T}$. V satisfies the definition of unfounded set for SDL with respect to D and $\mathcal{I}$. □

Lemma 6.5 (T and U are monotone)**.** *:If $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$, then $T_D(\mathcal{I}_1) \subseteq T_D(\mathcal{I}_2)$, and $U_D(\mathcal{I}_1) \subseteq U_D(\mathcal{I}_2)$.*

Proof. Suppose $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$ and let $p \in T_D(\mathcal{I}_1)$. Then either there exists an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_1$, or there exists a rule $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_1$ and for each conflict set $c \in C[p]$, $c - \{p\} \cap \mathcal{U}_1 \neq \varnothing$. Since $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$, we have $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{U}_1 \subseteq \mathcal{U}_2$. Thus either there exists an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_2$, or there exists a rule $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_2$ and for each conflict set $c \in C[p]$, $c - \{p\} \cap \mathcal{U}_2 \neq \varnothing$. It can be seen that $p \in T_D(\mathcal{I}_2)$.

Suppose now that $p \in U_D(\mathcal{I}_1)$. By definition, $U_D(\mathcal{I}_1)$ is the greatest unfounded set *wrt* D and $\mathcal{I}_1$, and so

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U}_1 \cup U_D(\mathcal{I}_1)) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

  (2.1) $body(r) \cap (\mathcal{U}_1 \cup U_D(\mathcal{I}_1)) \neq \varnothing$, or
  (2.2) there is a $c \in C[p]$ such that $c - \{p\} \subseteq \mathcal{T}_1$.

As before, $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{U}_1 \subseteq \mathcal{U}_2$ and so substitution yields

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U}_2 \cup U_D(\mathcal{I}_1)) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

  (2.1) $body(r) \cap (\mathcal{U}_2 \cup U_D(\mathcal{I}_1)) \neq \varnothing$, or
  (2.2) there is a $c \in C[p]$ such that $c - \{p\} \subseteq \mathcal{T}_2$.

Generalizing on p, it can be seen that $U_D(\mathcal{I}_1)$ is an unfounded set *wrt* D and $\mathcal{I}_2$. Since by definition $U_D(\mathcal{I}_2)$ is the union of all unfounded sets *wrt* D and $\mathcal{I}_2$, it follows that $p \in U_D(\mathcal{I}_2)$. □

Lemma 6.8 ($\mathcal{I}$ is monotone nondecreasing)**.** *Let D be a defeasible theory. For any $\lambda \geq 0$, $\mathcal{I}_{D,\lambda} \sqsubseteq \mathcal{I}_{D,\lambda+1}$.*

Proof. This follows immediately from the definition of the sequence $\mathcal{I}_D$ and the monotonicity of T_D and U_D. □

Lemma 6.9 (Each $\mathcal{I}_\lambda \in \mathcal{I}$ is coherent)**.** *For any $\lambda \geq 0$, if $p \in \mathcal{T}_\lambda$, then $p \notin \mathcal{U}_\lambda$.*

Proof. For $\lambda = 0$, the lemma holds. Suppose it holds for all $\alpha \leq \lambda$ and let λ be a successor ordinal. Let A be some set of literals such that $A \cap \mathcal{T}_\lambda \neq \varnothing$. Since $A \cap \mathcal{T}_\lambda \neq \varnothing$, there must be some least successor ordinal $\kappa \leq \lambda$ such that $\mathcal{T}_\kappa \cap A \neq \varnothing$ and for all $\eta < \kappa$, $\mathcal{T}_\eta \cap A = \varnothing$.

Let p be a literal such that $p \in A$ and $p \in \mathcal{T}_\kappa$. Then either definition 6.3.1 or 6.3.2 holds (recall that 6.3 is the definition of immediate consequence for SDL). Suppose it's 6.3.1. Then there is an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_{\kappa-1}$. By choice of A, $A \cap \mathcal{T}_{\kappa-1} = \varnothing$. So, $body(r) \cap A = \varnothing$. By Lemma 6.8 above, since $body(r) \subseteq \mathcal{T}_{\kappa-1}$, $body(r) \subseteq \mathcal{T}_{\lambda-1}$. By inductive hypothesis, $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$. Since $p \in A$ and there exists an $r \in R_s[p]$ such that $body(r) \cap (\mathcal{U}_\lambda \cup A) = \varnothing$, A fails the definition of unfound set with respect to $\mathcal{I}_{\lambda-1}$.

Suppose it's 6.3.2. Then there is an $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_{\kappa-1}$ and for each conflict set $c \in C[p]$, there exists a $q \in c - \{p\}$ such that $q \in \mathcal{U}_{\kappa-1}$. By choice of A, $A \cap \mathcal{T}_{\kappa-1} = \varnothing$ and so $body(r) \cap A = \varnothing$. By monotonicity of $\mathcal{I}$, since $body(r) \subseteq \mathcal{T}_{\kappa-1}$, it follows that $body(r) \subseteq \mathcal{T}_{\lambda-1}$ and so by the inductive hypothesis $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$. Also by monotonicity of $\mathcal{I}$, if $q \in \mathcal{U}_{\kappa-1}$ then $q \in \mathcal{U}_{\lambda-1}$. By the inductive hypothesis, $q \notin \mathcal{T}_{\lambda-1}$. Thus there exists a $p \in A$ and a rule $r \in R_d[p]$ such that $body(r) \cap (\mathcal{U}_{\lambda-1} \cup A) = \varnothing$ and for all conflict sets $c \in C[p]$ there is a q such that $q \notin \mathcal{T}_{\lambda-1}$. A again violates the definition of unfounded set with respect to D and $\mathcal{I}_\lambda$.

Generalizing on A, no set that intersects $\mathcal{T}_\lambda$ is unfounded with respect to D and $\mathcal{I}_{\lambda-1}$. In other words, if a set A is unfounded *wrt* D and $\mathcal{I}_{\lambda-1}$, then it does not intersect $\mathcal{T}_\lambda$. In particular, $\mathcal{U}_\lambda$—the greatest unfounded set with respect to D and $\mathcal{I}_{\lambda-1}$—does not intersect $\mathcal{T}_\lambda$. We conclude that for any $p \in \mathcal{T}_\lambda$, $p \notin \mathcal{U}_\lambda$.

(Limit Case) Suppose λ is a limit ordinal and the hypothesis holds for all $\alpha < \lambda$. If $\mathcal{T}_\lambda \cap \mathcal{U}_\lambda \neq \varnothing$, then for some p we have $p \in \mathcal{T}_\lambda$ and $p \in \mathcal{U}_\lambda$. By definition of $\mathcal{I}_\lambda$, there must be a successor ordinal $\kappa < \lambda$ such that $p \in \mathcal{T}_\kappa$ and $p \in \mathcal{U}_\kappa$. This contradicts the assumption that the hypothesis holds for all $\alpha < \lambda$, and so $p \notin \mathcal{U}_\lambda$. □

D.2 Soundness and Completeness

Theorem 6.14 (Completeness for Finite Grounded Components)**.** *Let D be a defeasible theory and $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, . . ., the sequence of interpretations created by iterating W_D from $\langle \varnothing, \varnothing \rangle$. For any $\alpha \geq 0$, If $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then*

(1) if $p \in \mathcal{T}_\mathcal{X}$, then $D \mid\sim_{SDL} p$, and
(2) if $p \in \mathcal{U}_\mathcal{X}$, then $D \sim\!\!|_{SDL} p$.

Proof. The proof is by induction on the sequence $(\mathcal{I}_D)$. $\mathcal{I}_0$ is empty and so the hypothesis trivially holds for $\alpha = 0$. Suppose the hypothesis holds for all $\lambda < \alpha$ and suppose α is a successor ordinal. Let $\mathcal{X}$ be a finite grounded component of $\mathcal{I}_{D,\alpha}$

(Case 1) Suppose $p \in \mathcal{T}_\mathcal{X}$. Then there exists a $\mathcal{Y}$ that is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$ and $p \in T_D(\mathcal{Y})$. As such there is some rule $r \in R_{sd}[p]$ such that $body(r) \subseteq \mathcal{T}_\mathcal{Y}$. If $r \in R_s[p]$, then by

inductive hypothesis $D \mid\sim_{SDL} a$ for each $a \in body(r)$, and so for each there exists a defeasible proof tree with root labeled $+\delta a$. We may append these proofs to a node labeled $+\delta p$ to form a proof showing $D \mid\sim_{SDL} p$. Since $body(r)$ is finite, the proof tree is finite.

If $r \in R_d[p]$, then as before for each a in $body(r)$, $D \mid\sim_{SDL} a$. Since $p \in T_D(\mathcal{Y})$, for all $c \in C[p]$ there is a $q \in c - \{p\}$ such that $q \in \mathcal{U}_\mathcal{Y}$. By inductive hypothesis, $D \sim|_{SDL} q$. Since $\mathcal{U}_\mathcal{Y}$ is finite, the number of q's is finite. Adding a tree for q for each conflict set $c \in C[p]$ as well as adding trees for each $a \in body(r)$ to a root labeled $+\delta p$ forms a proof tree that satisfies definition 6.12. And so $D \mid\sim_{SDL} p$.

(Case 2) Suppose $p \in \mathcal{U}_\mathcal{X}$. Then there exists a $\mathcal{Y} \sqsubseteq \mathcal{I}_{D,\alpha-1}$ such that $\mathcal{Y}$ is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$, and furthermore, $\mathcal{U}_\mathcal{X}$ is an unfounded set *wrt* D and $\mathcal{Y}$.

Let τ_0 be the tree consisting of a single unmarked node labeled $-\delta p$. From τ_0, we construct a series of trees. Given a tree τ_i, we form a new tree τ_{i+1} by picking any unmarked node $x \in \tau_i$ labeled $-\delta q$ for some q such that $q \in \mathcal{U}_\mathcal{X}$. Since $q \in \mathcal{U}_\mathcal{X}$ and $\mathcal{U}_\mathcal{X}$ is an unfounded set *wrt* D and $\mathcal{Y}$, for each rule $r \in R_{sd}[q]$, there is literal $a \in body(r)$ such that either (a) $a \in \mathcal{U}_\mathcal{Y}$ or (b) $a \in \mathcal{U}_\mathcal{X}$ or (c) $r \in R_d$ and there exists a conflict set $c \in C[q]$ such that $c - \{q\} \subseteq \mathcal{T}_\mathcal{Y}$. We consider each rule $r \in R[q]$, treating the cases (a), (b), and (c) in turn.

(2.a) $a \in \mathcal{U}_\mathcal{Y}$. By inductive hypothesis $D \sim|_{SDL} a$. We may append a proof tree τ_a with root $-\delta a$ to node x and mark each node of τ_a.

(2.b) $a \in \mathcal{U}_\mathcal{X}$. If x does not already have a child labeled $-\delta a$, then append to x a node y labeled $-\delta a$. If y satisfies condition 2 or 3 in Definition 6.12, then mark y. Otherwise, leave y unmarked.

(2.c) Since $c - \{q\} \subseteq \mathcal{T}_\mathcal{Y}$, then by inductive hypothesis proof trees exist for each literal in $c - \{p\}$. Append these trees to x and mark every node occurring in them.

After applying one of the cases 2.a–2.c for each rule $r \in R_{sd}[q]$, examine the resulting tree to see if there is an unmarked non-leaf node z in the tree such that all the children of z are marked. If

such a node z is found, mark it. Repeat this procedure until there are no more unmarked nodes in the tree all of whose children are marked. The resulting tree is τ_{i+1}.

$$\text{Let } \tau = \bigcup_{i=0}^{\infty} \tau_i.$$

Suppose x is a marked node in τ. If x was added to τ using case 2.a, then x occurs within a subtree of τ that is a proof tree. So x must satisfy one of the conditions in Definition 6.12. Similarly, if x was added using case 2.c, then x is part of a valid proof tree and so satisfies the proof conditions. If x was added to τ and marked according to case 2.b, then x is a leaf node in τ and x satisfies condition 2 or 3 of Definition 6.12. Otherwise, x is a non-leaf node in τ, x was added to τ using condition 2.b, and x was marked because all of its children were marked. Looking at the cases used to add the children of x to τ (we add a child for each rule for q), we see that x must satisfy condition 2 in Definition 6.12. So if τ is finite and if every node in τ is marked, then τ is a proof tree.

Since cases 2.a–2.c append to node x nodes labeled with a literal from $\mathcal{X}$ or $\mathcal{Y}$ and both of these are finite, it must be the case that the branching factor of τ is finite. So if τ is infinite, then τ must have an infinitely long branch. Consider such a branch. Every node in this branch (other than the top node) must have been added using case 2.b since all the other branches add proof trees which are finite. So every node in the branch must be labeled $-\delta q$ for some literal q. Furthermore, no node in the branch satisfies condition 3 in Definition 6.12 since if it did, it would have been marked when it was added to τ and it would therefore have no children. But since $\mathcal{X}$ is finite, only finitely many literals occur in $\mathcal{X}$. So there must be some literal q such that two different nodes in our infinite branch are labeled $-\delta q$. But then one of these two nodes does satisfy condition 3 of Definition 6.12, which is a contradiction. Therefore, τ is not infinite.

Since τ is not infinite, we can let j be a non-negative integer such that $\tau = \tau_j$. Suppose τ_j has an unmarked node. Since a node must be marked if all its children are marked, τ_j must have an unmarked leaf node x. This node must have been added by case 2.b of our construction, and so we can let q be a literal such that x is labeled $-\delta q$, and $q \in \mathcal{U}_{\mathcal{X}}$. Since x is not marked, it satisfies

neither condition 2 or 3 of Definition 6.12. If there is no rule $r \in R_{sd}[q]$, then x satisfies condition 2 of Definition 6.12. So there is a rule $r \in R_{sd}[q]$, and one of the cases 2.a–2.c applies to x. So there must be some $m > j$ such x has a child node in τ_m. Then x is not a leaf node in τ_m and x is not a leaf node in τ, a contradiction. Therefore, every node in τ satisfies some condition in Definition 6.12 and τ is a proof tree.

If α is a limit ordinal, then by Lemma 4.21, there is a $\beta < \alpha$ such that $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\beta}$. By inductive hypothesis, if $p \in \mathcal{T}_{\mathcal{X}}$, then $D \mid\!\sim_{SDL} p$, and if $p \in \mathcal{U}_{\mathcal{X}}$, then $D \sim\!\mid_{SDL} p$. □

Theorem 6.13 (Soundness)**.** *Let D be a defeasible theory and $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, ..., the sequence of interpretations created by iterating W_D from $\langle \varnothing, \varnothing \rangle$.*

(1) If $D \mid\!\sim_{SDL} p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_{\mathcal{X}}$.

(2) If $D \sim\!\mid_{SDL} p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{U}_{\mathcal{X}}$.

Proof. If $D \mid\!\sim_{SDL} p$ or $D \sim\!\mid_{SDL} p$, then there is a proof tree τ showing this. We induct on the depth of τ.

(Base case) Suppose τ is just a single node n labeled $+\delta p$ or $-\delta p$. We consider each case separately.

(Case 1) Suppose that n is labeled $+\delta p$. Then either 6.12.1.a or 6.12.1.b of the definition of proof obtains. If it's 6.12.1.a, since n has no children, there must be a rule $r \in R_s[p]$ such that $body(r) = \varnothing$. By definition 6.3 (Immediate Consequence), $p \in T_D(\mathcal{I}_0)$. But $T_D(\mathcal{I}_0) = \mathcal{T}_1$, and so $p \in \mathcal{T}_1$. Since $\{\}$ is unfounded *wrt* D and $\mathcal{I}_0$ and $p \in T_D(\mathcal{I}_0)$, the interpretation $\langle \{p\}, \{\} \rangle$ forms a finite grounded component of $\mathcal{I}_1$.

If 6.12.1.b obtains, then there is some rule $r \in R_d[p]$ that succeeds at n. Since n has no children, $body(r) = \varnothing$. Let $c \in C[p]$. Since definition 6.12.1.b is satisfied, $c - \{p\}$ fails at n. Since τ consists of a single node, no set can fail at n, and so there can be no conflict set $c \in C[p]$. By definition

6.3.2, $p \in T_D(\mathcal{I}_0) = \mathcal{T}_1$. As before, the interpretation $\langle\{p\}, \{\}\rangle$ forms a finite grounded component of $\mathcal{I}_1$.

(Case 2) Suppose that n is labeled $-\delta p$. Since τ consists of only a single node, failure-by-looping cannot apply and also there can be no strict rules with head p. Therefore 6.12.2.b must obtain. Suppose that there exists an $r \in R_d[p]$. Since node n has no children, $body(r) = \varnothing$ and so 6.12.2.b.ii must hold. Thus there is a conflict set $c \in C[p]$ such that $c - \{p\}$ succeeds. Since τ has only a single node, $c - \{p\}$ must be $\{\}$. Observe that $\{\} \subseteq \mathcal{T}_0$. Since there are no strict rules for p and for every defeasible rule $r \in R_d[p]$, there exists a conflict set $c \in C[p]$ such that $c - \{p\} \subseteq \mathcal{T}_0$, $\{p\}$ is an unfounded set *wrt* D and $\mathcal{I}_0$. The interpretation $\langle\{\}, \{p\}\rangle$ thus forms a finite grounded component of $\mathcal{I}_1$.

(Induction) Suppose for all $j \leq k$, if τ has depth j and its root is labeled $+\delta p$ ($-\delta p$), then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_{\mathcal{X}}$ ($p \in \mathcal{U}_{\mathcal{X}}$). Suppose τ has depth $k + 1$.

(Case 1) Suppose the root n of τ is labeled $+\delta p$. Then 6.12.1.a or 6.12.1.b again holds. If 6.12.1.a holds, then there is a strict rule $r \in R_s[p]$ such that $body(r)$ succeeds at n. For all $q \in body(r)$, there is a child of m labeled $+\delta q$. Each such q is the root of a valid argument tree of maximum depth k, and so by inductive hypothesis, there exists a finite α_q and a finite grounded component $\mathcal{X}_q$ of $\mathcal{I}_{D,\alpha_q}$ such that $q \in \mathcal{T}_{\mathcal{X}_q}$. From Lemma 4.20, there exists a least $\beta \geq 0$ such that for each $q \in body(r)$, $\mathcal{X}_q \sqsubseteq \mathcal{I}_{D,\beta}$. Since each α_q is finite, β must be finite as well. Let $\mathcal{X}$ be the least upper bound of these $\mathcal{X}_q$'s. Since $body(r)$ is finite, $\mathcal{X}$ is finite. From Lemma 4.22, $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\beta}$. Thus, we have $body(r) \subseteq \mathcal{T}_{\mathcal{X}}$ and $r \in R_s$. By Definition 6.3.1, $p \in T_D(\mathcal{X})$. Since $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\beta}$ and $p \in T_D(\mathcal{X})$, by monotonicity of T_D, $p \in T_D(\mathcal{I}_{D,\beta})$. The interpretation $\langle\{p\}, \{\}\rangle$ thus forms a finite grounded component of $\mathcal{I}_{\beta+1}$.

If 6.12.1.b holds, then there is a defeasible rule $r \in R_d[p]$ such that $body(r)$ succeeds at n. As before, for each $q \in body(r)$, we have a finite α_q and finite grounded component $\mathcal{X}_q$ of $\mathcal{I}_{D,\alpha_q}$ such that $q \in \mathcal{T}_{\mathcal{X}_q}$. Let $c \in C[p]$. Since 6.12.1.b holds, there is a $u \in c - \{p\}$ such that$\{u\}$ fails at n. By definition of failure, there exists a child m of n labeled $-\delta u$. Node m is thus the head of

a valid proof tree of depth $\leq k$, and so by inductive hypothesis, there exists a finite α_u and finite grounded component $\mathcal{X}_u$ of $\mathcal{I}_{D,\alpha_u}$ such that $u \in \mathcal{U}_{\mathcal{X}_u}$. Generalizing on c, for every $c \in C[p]$, there is a $u \in c - \{p\}$ such that $u \in \mathcal{U}_{\mathcal{X}_u}$ for some finite grounded component $\mathcal{X}_u$ of $\mathcal{I}_{D,\alpha_u}$.

From Lemmas 4.20, there exists a least $\beta \geq 0$ such that each $\mathcal{X}_q \sqsubseteq \mathcal{I}_{D,\beta}$ and $\mathcal{X}_u \sqsubseteq \mathcal{I}_{D,\beta}$. Let $\mathcal{X}$ be the least upper bound of these $\mathcal{X}_q$'s and $\mathcal{X}_u$'s. Since τ is finite, $\mathcal{X}$ is finite. Also, since each α is finite, β is finite. From Lemma 4.22, $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\beta}$. Thus, $body(r) \subseteq \mathcal{T}_{\mathcal{X}}$ and for every $c \in C[p]$, there is a $q \in c - \{p\}$ such that $q \in \mathcal{U}_{\mathcal{X}}$. By definition of 6.3.2, $p \in T_D(\mathcal{X})$. The interpretation $\langle\{p\}, \{\}\rangle$ again forms a finite grounded component of $\mathcal{I}_{\beta+1}$.
(Case 2) Suppose the root n of τ is labeled $-\delta p$. Any branch of a proof tree involving failure-by-looping need not extend beyond the topmost node where definition 6.12.3 (failure-by-looping) applies. As this is so, the tree can be trimmed to that point, and so 6.12.3 only applies to the leaves of the tree. We may assume without loss of generality that τ is of this form.

Define N to be the set of nodes of τ labeled with $-\delta u$ for any u, and S to be the set of the u's. Let n be any node in N. Then n is labeled $-\delta q$ for some q. Node n is either a leaf or an internal node. We treat each case separately.
(Case 2.a) If n is an internal node, then 6.12.2 obtains. Suppose $r \in R_{sd}[q]$. If $r \in R_s[q]$, then $body(r)$ fails at n. By definition of failure, n has a child m labeled $-\delta v$, where $v \in body(r)$. By definition of N and S, $m \in N$ and $v \in S$.

Let $r \in R_d[q]$. Since 6.12.2 holds at n, either (i) the $body(r)$ fails at n and so there is a $a \in body(r)$ and a child m of n such that m is labeled $-\delta a$ and $m \in N$ (and so $a \in S$), or (ii) there is a conflict set $c \in C[q]$ such that for all $u \in c - \{q\}$, there is a child m of n labeled $+\delta u$. If (ii) holds, then there is a subtree of n with root labeled $+\delta u$ that constitutes a valid proof. This subtree has depth $\leq k$ and so by inductive hypothesis there exists a finite $\alpha_u \geq 0$ and finite grounded component $\mathcal{X}_u$ of $\mathcal{I}_{D,\alpha_u}$ such that $u \in \mathcal{T}_{\mathcal{X}_u}$. Generalizing on u, since τ is finite, then by Lemmas 4.20 and 4.22, there exists a least ordinal β and single finite grounded component $\mathcal{X}_r$ of $\mathcal{I}_{D,\beta}$ such that $c - \{q\} \subseteq \mathcal{T}_{\mathcal{X}_r}$.

Observe that there is a $\mathcal{X}_r$ for each rule $r \in R_{sd}[q]$. $R_{sd}[q]$ might be infinite in size, but since τ is finite, we are assured that the number of $\mathcal{X}_r$'s is finite. Each $\mathcal{X}_r$ is a finite grounded component of some $\mathcal{I}_{D,\beta}$. Let λ be the maximum of the β's and $\mathcal{X}$ the least upper bound of the $\mathcal{X}_r$'s. From Lemmas 4.20 and 4.22 $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\lambda}$. Thus, generalizing on r, (1) for each $r \in R_s[q]$, $body(r) \cap S \neq \varnothing$, and (2) for each $r \in R_d[q]$ either $body(r) \cap S \neq \varnothing$, or else there is a conflict set $c \in C[q]$ such that $c - \{q\} \subseteq \mathcal{T}_{\mathcal{X}}$.

(Case 2.b) Suppose that n is a leaf node. Then either 6.12.2 or 6.12.3 obtains. If 6.12.2 obtains, then as was shown in the base case, $\langle\{\}, \{q\}\rangle$ is a finite grounded component of $\mathcal{I}_{D,1}$. If 6.12.3 obtains, then there is a non-leaf node labeled $-\delta q$, and we have shown there that (1) for each $r \in R_s[q]$, $body(r) \cap S \neq \varnothing$, and (2) for each $r \in R_d[q]$ either $body(r) \cap S \neq \varnothing$, or else there is a conflict set $c \in C[q]$ such that $c - \{q\} \subseteq \mathcal{T}_{\mathcal{X}}$.

Given the above 2 cases, by definition S is unfounded with respect to D and $\mathcal{X}$. Note that S is a finite set. As this is so, $\langle\{\}, S\rangle$ forms a finite grounded component with respect to $\lambda + 1$. $\square$

APPENDIX E

PROOFS OF THEOREMS FOR MDL

We now prove monotonicity and coherence of the operators defined for MDL as well as soundness and completeness of its proof system *wrt* the semantics. In the following, the operators T_D, U_D, and W_D are assumed to be defined according to MDL.

E.1 MONOTONICITY, COHERENCE

Lemma E.1 (Unfounded sets are closed under union)**.** *If S_0, S_2, ..., be unfounded sets of MDL* wrt *to defeasible theory D and interpretation $\mathcal{I}$, then $\bigcup S_i$ is unfounded* wrt *to D and $\mathcal{I} = \langle \mathcal{T}, \mathcal{U} \rangle$.*

Proof. Let $V = \bigcup_{i=0}^{\infty} S_i$. Without loss of generality, suppose that $p \in S_k$ for some k. If $r \in R_s[p]$, then since S_k is unfounded, $body(r) \cap (S_k \cup \mathcal{U}) \neq \varnothing$. But since $S_k \subseteq V$, $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$. Similarly, if $r \in R_d[p]$, then either $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$ or there is a $c \in C[r]$ such that for each $s \in c - \{r\}$, $\{head(s)\} \cup body(s) \subseteq \mathcal{T}$ and $s \not\prec r$. Generalizing on r and p, for all $p \in V$, if $r \in R_s[p]$, $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$. If $r \in R_d[p]$, then either $body(r) \cap (V \cup \mathcal{U}) \neq \varnothing$ or there is a $c \in C[r]$ such that for all $s \in c - \{r\}$, $\{head(s)\} \cup body(s) \subseteq \mathcal{T}$ and $s \not\prec r$. V satisfies the definition of unfounded set for SDL with respect to D and $\mathcal{I}$. □

Lemma E.2 (T and U are monotone)**.** *: If $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$, then $T_D(\mathcal{I}_1) \subseteq T_D(\mathcal{I}_2)$, and $U_D(\mathcal{I}_1) \subseteq U_D(\mathcal{I}_2)$.*

Proof. Suppose $\mathcal{I}_1 \sqsubseteq \mathcal{I}_2$ and let $p \in T_D(\mathcal{I}_1)$. Then either there exists an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_1$, or there exists a rule $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_1$ and for each conflict set $c \in C[r]$, there exists a $s \in c - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_1 \neq \varnothing$ or $s \prec r$. Since $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{U}_1 \subseteq \mathcal{U}_2$, either there exists an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_2$, or there exists a

rule $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_2$ and for each conflict set $c \in C[r]$, there exists a $s \in c - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_2 \neq \varnothing$ or $s \prec r$. By definition of T_D, $p \in T_D(\mathcal{I}_2)$.

Now suppose $p \in U_D(\mathcal{I}_1)$. $U_D(\mathcal{I}_1)$ is by definition the greatest unfounded set *wrt* $\mathcal{I}_1$, and so

(1) For every $r \in R_s[p]$, $body(r) \cap (\mathcal{U}_1 \cup U_D(\mathcal{I}_1)) \neq \varnothing$.
(2) For every $r \in R_d[p]$,

(2.1) $body(r) \cap (\mathcal{U}_1 \cup U_D(\mathcal{I}_1)) \neq \varnothing$, or
(2.2) there is a $c \in C[r]$ such that for each $s \in c - \{r\}$, $\{head(s)\} \cup body(s) \in \mathcal{T}_1$ and $s \not\prec r$.

As before, we may substitute $\mathcal{T}_2$ for $\mathcal{T}_1$ and $\mathcal{U}_2$ for $\mathcal{U}_1$ in the above conditions. Generalizing on p, it can be seen that $U_D(\mathcal{I}_1)$ is unfounded *wrt* D and $\mathcal{I}_2$. By definition, $U_D(\mathcal{I}_2)$ is the union of all unfounded sets *wrt* D and $\mathcal{I}_2$. And so $p \in U_D(\mathcal{I}_2)$. □

Lemma E.3 ($(\mathcal{I})$ is monotone nondecreasing)**.** *For any $\lambda \geq 0$, if $p \in \mathcal{T}_\lambda$, then $p \in \mathcal{T}_{\lambda+1}$, and if $p \in \mathcal{U}_\lambda$, then $p \in \mathcal{U}_{\lambda+1}$.*

Proof. Immediate from the definitions and the above Lemma. □

The sequence of interpretations defined using W_D is coherent, in the sense that for any literal p, p cannot both be in $\mathcal{T}$ and $\mathcal{U}$.

Lemma E.4 ($(\mathcal{I})$ is coherent)**.** *For any $\lambda \geq 0$, if $\mathcal{T}_\lambda \cap \mathcal{U}_\lambda = \varnothing$.*

Proof. For $\lambda = 0$, the lemma holds. Suppose it holds for all $\alpha < \lambda$. Let A be some set of literals such that $A \cap \mathcal{T}_\lambda \neq \varnothing$. Since $A \cap \mathcal{T}_\lambda \neq \varnothing$, there must be some least $\kappa \leq \lambda$ such that $\mathcal{T}_\kappa \cap A \neq \varnothing$ and for all $\eta < \kappa$, $\mathcal{T}_\eta \cap A = \varnothing$.

Let p be a literal such that $p \in A$ and $p \in \mathcal{T}_\kappa$. Then either definition 6.42.1 or 6.42.2 holds. Suppose it's 6.42.1. Then there is an $r \in R_s[p]$ such that $body(r) \subseteq \mathcal{T}_{\kappa-1}$. By choice of A, $A \cap \mathcal{T}_{\kappa-1} = \varnothing$. So, $body(r) \cap A = \varnothing$. By monotonicity of T_D, since $body(r) \subseteq \mathcal{T}_{\kappa-1}$, $body(r) \subseteq \mathcal{T}_{\lambda-1}$. By inductive hypothesis, $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$. Since $p \in A$ and there exists an $r \in R_s[p]$ such that $body(r) \cap (\mathcal{U}_{\lambda-1} \cup A) = \varnothing$, A fails the definition of unfound set with respect to $\mathcal{I}_{\lambda-1}$.

Suppose it's 6.42.2. Then there is an $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_{\kappa-1}$ and for each conflict set $c \in C[r]$, there exists an $r_q \in c - \{r\}$ such that $(\{head(r_q)\} \cup body(r_q)) \cap \mathcal{U}_{\kappa-1} \neq \varnothing$ or $r_q \prec r$.

By choice of A, $A \cap \mathcal{T}_{\kappa-1} = \varnothing$ and so $body(r) \cap A = \varnothing$. By monotonicity of $\mathcal{I}$, since $body(r) \subseteq \mathcal{T}_{\kappa-1}$, it follows that $body(r) \subseteq \mathcal{T}_{\lambda-1}$ and so by inductive hypothesis $body(r) \cap \mathcal{U}_{\lambda-1} = \varnothing$. Also by monotonicity of $\mathcal{I}$, if $(\{head(r_q)\} \cup body(r_q)) \cap \mathcal{U}_{\kappa-1} \neq \varnothing$ then $(\{head(r_q)\} \cup body(r_q)) \cap \mathcal{U}_{\lambda-1} \neq \varnothing$. Since $\mathcal{U}_{\lambda-1}$ and $\mathcal{T}_{\lambda-1}$ are disjoint, we have $(\{head(r_q) \cup body(r_q)) \nsubseteq \mathcal{T}_{\lambda-1}$ or $r_q \prec r$. In summary, there exists an $r \in R_d[p]$ such that $body(r) \subseteq \mathcal{T}_{\lambda-1}$ and for each conflict set $c \in C[r]$, there exists an $r_q \in c - \{r\}$ for which it is not the case that $\{headr_q \cup body(r_1) \subseteq \mathcal{T}_{\lambda-1}$ and $r_q \not\prec r$. A again violates the definition of unfounded set with respect to $\mathcal{I}_{\lambda-1}$.

Generalizing on A, no set A that intersects $\mathcal{T}_\lambda$ is unfounded with respect to $\mathcal{I}_{\lambda-1}$. In other words, if a set A is unfounded *wrt* $\mathcal{I}_{\lambda-1}$, then it does not intersect $\mathcal{T}_\lambda$. In particular, $\mathcal{U}_\lambda$—the greatest unfounded set with respect to $\mathcal{I}_{\lambda-1}$—does not intersect $\mathcal{T}_\lambda$. We conclude that for any $p \in \mathcal{T}_\lambda$, $p \notin \mathcal{U}_\lambda$.

If λ is a limit ordinal, suppose that $\mathcal{T}_\lambda \cap \mathcal{U}_\lambda \neq \varnothing$. By definition of $\mathcal{I}_\lambda$ there must be a least successor ordinal $\kappa < \lambda$ such that $\mathcal{T}_\kappa \cap \mathcal{U}_\kappa \neq \varnothing$. But this violates the inductive hypothesis. And so it must be that $\mathcal{T}_\lambda \cap \mathcal{U}_\lambda \neq \varnothing$. □

E.2 Soundness and Completeness

Theorem 6.50 (Soundness)**.** *Let D be a defeasible theory. If $D \mid\sim_{MDL} p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_\mathcal{X}$. If $D \sim\!\!\mid_{MDL} p$, then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{U}_\mathcal{X}$.*

Proof. If $D \mid\sim_L p$ or $D \sim\!\!\mid_L p$, then there is a proof tree τ showing this. We induct on the depth of τ.

(base case) Suppose τ is just a single node n labeled $+\delta p$ or $-\delta p$. We consider each case separately.

(Case 1) Suppose that n is labeled $+\delta p$. Then either 6.49.1.1 or 6.49.1.2 obtains. If it's 6.49.1.1, since n has no children, there must be a rule $\{\} \rightarrow p \in R_D$. By definition 6.42.1 (the immediate consequence operator), $p \in T_D(\mathcal{I}_0)$. But $T_D(\mathcal{I}_0) = \mathcal{T}_1$, and so $p \in \mathcal{T}_1$. Since $\{\}$ is unfounded *wrt* D and $\mathcal{I}_0$ and $p \in T_D(\mathcal{I}_0)$, the interpretation $\langle\{p\}, \{\}\rangle$ forms a finite grounded component of $\mathcal{I}_1$.

If 6.49.1.2 obtains, then there is some rule $r \in R_d[p]$ that $body(r)$ succeeds at n. Since n has no children, $body(r) = \varnothing$. Let $c \in C[r]$. Since definition 6.49.1.2 is satisfied, then there is some $s \in c - \{r\}$ such that $\{head(s)\} \cup body(s)$ fails at n or else $s \prec r$. Since τ consists of a single node, no set can fail at n, and so it must be the case that $s \prec r$. Generalizing on c, each conflict set $c \in C[r]$ contains a $s \in c - \{r\}$ such that $s \prec r$. By definition 6.42.2, $p \in T_D(\mathcal{I}_0) = \mathcal{T}_1$. As before, the interpretation $\langle\{p\}, \{\}\rangle$ forms a finite grounded component of $\mathcal{I}_1$.

(Case 2) Suppose that n is labeled $-\delta p$. Since τ consists of only a single node, failure-by-looping cannot apply and there can be no strict rules with head p. Therefore 6.49.2.2 must obtain. Suppose that there exists an $r \in R_d[p]$. Since node n has no children, $body(r)$ cannot fail at n and so 6.49.2.2.2 must hold. Thus there is a conflict set $c \in C[r]$ such that for all $s \in c - \{r\}$, there is a child of n labeled $+\delta head(s)$. Since τ has only a single node, $c - \{r\}$ must be $\{\}$. Observe that $\{\} \subseteq \mathcal{T}_0$. Since there are no strict rules for p and for every defeasible rule $r \in R_d[p]$, there exists a conflict set $c \in C[r]$ such that for each $s \in c - \{r\}$ we have $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_0$ and $s \not\prec r$, $\{p\}$ is an unfounded set *wrt* D and $\mathcal{I}_0$. The interpretation $\langle\{\}, \{p\}\rangle$ thus forms a finite grounded component of $\mathcal{I}_1$.

(Induction) Suppose for all $j \leq k$, if τ has depth j and its root is labeled $+\delta p$ ($-\delta p$), then there exists a finite $\alpha \geq 0$ and a finite grounded component $\mathcal{X}$ of $\mathcal{I}_{D,\alpha}$ such that $p \in \mathcal{T}_\mathcal{X}$ ($p \in \mathcal{U}_\mathcal{X}$). Suppose τ has depth $k + 1$.

(Case 1) Suppose the root n of τ is labeled $+\delta p$. Then 6.49.1.1 or 6.49.1.1 again holds. If 6.49.1.1 holds, then there is a strict rule $r \in R_s[p]$ such that $body(r)$ succeeds at n. For all $q \in body(r)$, there is a child of m labeled $+\delta q$. Each such q is the root of a valid argument tree of maximum depth k, and so by inductive hypothesis, there exists a finite α_q and a finite grounded component $\mathcal{X}_q$ of $\mathcal{I}_{D,\alpha_q}$ such that $q \in \mathcal{T}_{\mathcal{X}_q}$. From Lemma 4.20, there exists a least $\beta \geq 0$ such that for each $q \in body(r)$, $\mathcal{X}_q \sqsubseteq \mathcal{I}_{D,\beta}$. Since each α_q is finite, β must be finite as well. Let $\mathcal{X}$ be the least upper bound of these $\mathcal{X}_q$'s. Since $body(r)$ is finite, $\mathcal{X}$ is finite. From Lemma 4.22, $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\beta}$. Thus, we have $body(r) \subseteq \mathcal{T}_\mathcal{X}$ and $r \in R_s$. By Definition

6.3.1, $p \in T_D(\mathcal{X})$. Since $\mathcal{X} \sqsubseteq \mathcal{I}_{D,\beta}$ and $p \in T_D(\mathcal{X})$, by monotonicity of T_D, $p \in T_D(\mathcal{I}_{D,\beta})$. The interpretation $\langle\{p\}, \{\}\rangle$ thus forms a finite grounded component of $\mathcal{I}_{\beta+1}$.

If 6.49.1.2 holds, then there is a defeasible rule $r \in R_d[p]$ such that $body(r)$ succeeds at n. As before, for all $q \in body(r)$,we have a finite α_q and finite grounded component $\mathcal{X}_q$ of $\mathcal{I}_{D,\alpha_q}$ such that $q \in \mathcal{T}_{\mathcal{X}_q}$.

Let $c \in C[r]$. Since 6.49.1.2 holds, there is a $s \in c - \{r\}$ such that $\{head(s)\} \cup body(s)$ fails at n or else $s \prec r$. If $\{head(s)\} \cup body(s)$ fails at n, there exists a $u \in \{head(s)\} \cup body(s)$ and child m of n labeled $-\delta u$. Node m is thus the head of a valid proof tree of depth $\leq k$, and so by inductive hypothesis, there exists a finite α_u and finite grounded component $\mathcal{X}_u$ of $\mathcal{I}_{D,\alpha_u}$ such that $u \in \mathcal{U}_{\mathcal{X}_u}$. Generalizing on c, for every $c \in C[r]$, there is a $s \in c - \{r\}$ such that $\{head(s)\} \cup body(s) \cap \mathcal{U}_{\mathcal{X}_u} \neq \varnothing$ for some finite grounded component $\mathcal{X}_u$ of $\mathcal{I}_{D,\alpha_u}$, or else $s \prec r$.

From Lemmas 4.20, there exists a least $\beta \geq 0$ such that each $\mathcal{X}_q \sqsubseteq \mathcal{I}_{D,\beta}$ and $\mathcal{X}_u \sqsubseteq \mathcal{I}_{D,\beta}$. Let $\mathcal{X}$ be the least upper bound of these $\mathcal{X}_q$'s and $\mathcal{X}_u$'s. Since τ is finite, $\mathcal{X}$ is finite. Also, since each α is finite, β is finite. From Lemma 4.22, $\mathcal{X}$ is thus a finite grounded component of $\mathcal{I}_{D,\beta}$. Thus, $body(r) \subseteq \mathcal{T}_{\mathcal{X}}$ and for every $c \in C[r]$, there is a $s \in c - \{p\}$ such that $\{head(s)\} \cup body(s) \cap \mathcal{U}_{\mathcal{X}} \neq \varnothing$, or else $s \prec r$. By definition of 6.42.2, $p \in T_D(\mathcal{X})$. The interpretation $\langle\{p\}, \{\}\rangle$ again forms a finite grounded component of $\mathcal{I}_{\beta+1}$.

(Case 2) Suppose the root n of τ is labeled $-\delta p$. Any branch of a proof tree involving failure-by-looping need not extend beyond the topmost node where definition 6.49.3 (failure-by-looping) applies. As this is so, the tree can be trimmed to that point, and so 6.49.3 only applies to the leaves of the tree. We may assume without loss of generality that τ is of this form.

Define N to be the set of nodes of τ labeled with $-\delta u$ for any u, and S to be the set of the u's. Let n be any node in N. Then n is labeled $-\delta q$ for some q. Node n is either a leaf or an internal node. We treat each case separately.

(Case 2.a) If n is an internal node, then 6.49.2 obtains. Suppose $r \in R_{sd}[q]$. If $r \in R_s[q]$, then $body(r)$ fails at n. By definition of failure, n has a child m labeled $-\delta v$, where $v \in body(r)$. By definition of N and S, $m \in N$ and $v \in S$.

Let $r \in R_d[q]$. Since 6.12.2 holds at n, either (i) the $body(r)$ fails at n and so there is a $a \in body(r)$ and a child m of n such that m is labeled $-\delta a$ and $m \in N$ (and so $a \in S$), or

(ii) there is a conflict set $c \in C[r]$ such that for all $s \in c - \{r\}$, $s \not\prec r$ and for each $u \in (\{head(s)\} \cup body(s))$, $\{u\}$ succeeds at n. If (ii) holds, then there is a subtree of n with root labeled $+\delta u$ that constitutes a valid proof. This subtree has depth $\leq k$ and so by inductive hypothesis there exists a finite $\alpha_u \geq 0$ and finite grounded component $\mathcal{X}_u$ of $\mathcal{I}_{D,\alpha_u}$ such that $u \in \mathcal{T}_{\mathcal{X}_u}$. Generalizing on u, since τ is finite, then by Lemmas 4.20 and 4.22, there exists a least ordinal β and single finite grounded component $\mathcal{X}_r$ of $\mathcal{I}_{D,\beta}$ such that for each $s \in c - \{r\}$, $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{\mathcal{X}_r}$ and $s \not\prec r$.

Observe that there is a $\mathcal{X}_r$ for each rule $r \in R_{sd}[q]$. $R_{sd}[q]$ might be infinite in size, but since τ is finite, we are assured that the number of $\mathcal{X}_r$'s is finite. Each $\mathcal{X}_r$ is a finite grounded component of some $\mathcal{I}_{D,\beta}$. Let λ be the maximum of the β's and $\mathcal{X}$ the least upper bound of the $\mathcal{X}_r$'s. From Lemmas 4.20 and 4.22 $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\lambda}$. Thus, generalizing on r, (1) for each $r \in R_s[q]$, $body(r) \cap S \neq \varnothing$, and (2) for each $r \in R_d[q]$ either $body(r) \cap S \neq \varnothing$, or else there is a conflict set $c \in C[r]$ such that for all $s \in c - \{r\}$ we have $\{head(s)\} \cup body(s) \subseteq \mathcal{T}_{\mathcal{X}}$ and $s \not\prec r$.

(Case 2.b) Suppose that n is a leaf node. Then either 6.49.2 or 6.49.3 obtains. If 6.49.2 obtains, then as was shown in the base case, $\langle\{\}, \{q\}\rangle$ is a finite grounded component of $\mathcal{I}_{D,1}$. If 6.49.3 obtains, then there is a non-leaf node labeled $-\delta q$, and we have shown there that (1) for each $r \in R_s[q]$, $body(r) \cap S \neq \varnothing$, and (2) for each $r \in R_d[q]$ either $body(r) \cap S \neq \varnothing$, or else there is a conflict set $c \in C[r]$ such that for all $s \in c - \{r\}$ we have $\{head(s)\} \cup body(s) \subseteq \mathcal{T}_{\mathcal{X}}$ and $s \not\prec r$.

Given the above 2 cases, by definition S is unfounded with respect to D and $\mathcal{X}$. Note that S is a finite set. As this is so, $\langle\{\}, S\rangle$ forms a finite grounded component with respect to $\lambda + 1$. □

Theorem 6.51 (Completeness for Finite Grounded Components)**.** *Let D be a defeasible theory and $\mathcal{I}_{D,0}$, $\mathcal{I}_{D,1}$, ... the sequence of interpretations created by W_D for MDL. For any $\alpha \geq 0$, If $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\alpha}$, then (1) if $p \in \mathcal{T}_{\mathcal{X}}$, then $D \mid\sim_{MDL} p$, and (2) if $p \in \mathcal{U}_{\mathcal{X}}$, then $D \sim\!\!\mid_{MDL} p$.*

Proof. The proof is by induction on the sequence $(\mathcal{I}_D)$. $\mathcal{I}_0$ is empty and so the hypothesis trivially holds for $\alpha = 0$. Suppose the hypothesis holds for all $\lambda < \alpha$ and suppose α is a successor ordinal. Let $\mathcal{X}$ be a finite grounded component of $\mathcal{I}_{D,\alpha}$

(Case 1) Suppose $p \in \mathcal{T}_\mathcal{X}$. Then there exists a $\mathcal{Y}$ that is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$ and $p \in T_D(\mathcal{Y})$. As such there is some rule $r \in R_{sd}[p]$ such that $body(r) \subseteq \mathcal{T}_\mathcal{Y}$. If $r \in R_s[p]$, then by inductive hypothesis $D \mid\sim_{MDL} a$ for each $a \in body(r)$, and so for each there exists a defeasible proof tree with root labeled $+\delta a$. We may append these proofs to a node labeled $+\delta p$ to form a proof showing $D \mid\sim_{MDL} p$. Since $body(r)$ is finite, the proof tree is finite.

If $r \in R_d[p]$, then as before for each a in $body(r)$, $D \mid\sim_{MDL} a$. Since $p \in T_D(\mathcal{Y})$, for all $c \in C[r]$ there is a $s \in c - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_\mathcal{Y} \neq \varnothing$ or else $s \prec r$. If $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_\mathcal{Y} \neq \varnothing$, then by inductive hypothesis, there exists a $v \in (\{head(s)\} \cup body(s))$ such that $D \sim\!|_{MDL} v$. Adding the tree $\tau_{-\delta v}$ as well as adding trees for each $a \in body(r)$ to a root labeled $+\delta p$ forms a proof tree for $D \mid\sim_{MDL} p$ that satisfies definition 6.49.1.

(Case 2) Suppose $p \in \mathcal{U}_\mathcal{X}$. Then there exists a $\mathcal{Y} \sqsubseteq \mathcal{I}_{D,\alpha-1}$ such that $\mathcal{Y}$ is a finite grounded component of $\mathcal{I}_{D,\alpha-1}$, and furthermore, $\mathcal{U}_\mathcal{X}$ is an unfounded set *wrt* D and $\mathcal{Y}$.

Let τ_0 be the tree consisting of a single unmarked node labeled $-\delta p$. From τ_0, we construct a series of trees. Given a tree τ_i, we form a new tree τ_{i+1} by picking any unmarked node x labeled $-\delta q$ for some q such that $q \in \mathcal{U}_\mathcal{X}$. Since $q \in \mathcal{U}_\mathcal{X}$ and $\mathcal{U}_\mathcal{X}$ is an unfounded set *wrt* D and $\mathcal{Y}$, for each rule $r \in R[q]$, there is literal $a \in body(r)$ such that (a) $a \in \mathcal{U}_\mathcal{Y}$ or (b) $a \in \mathcal{U}_\mathcal{X}$, or else (c) r is defeasible and there is a conflict set $c \in C[r]$ such that for each $s \in c - \{r\}$, $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_\mathcal{Y}$ and $s \not\prec r$. We consider each rule $r \in R[q]$, treating the cases (a), (b), and (c) in turn.

(Case 2.a) $a \in \mathcal{U}_\mathcal{Y}$. By inductive hypothesis $D \sim\!|_{MDL} a$. We may append a proof tree for this to node x and mark each node of the appended subtree.

(Case 2.b) $a \in \mathcal{U}_\mathcal{X}$. If x does not already have a child labeled $-\delta a$, then append to x a node y labeled $-\delta a$. If y satisfies condition 2 or 3 in Definition 6.49, then mark y. Otherwise, leave y unmarked.

(Case 2.c) there is a conflict set $c \in C[r]$ such that for each $s \in c - \{r\}$, $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{\mathcal{Y}}$ and $s \not\prec r$. By inductive hypothesis, for each literal $v \in \{head(s)\} \cup body(s)$, we have $D \mid\!\sim_{MDL} v$. Append trees showing this to node x and mark every node occurring in them.

After applying one of the cases 2.a, 2.c, or 2.b for each rule r with head q, examine the resulting tree (including node x) to see if there is an unmarked non-leaf node z in the tree such that all the children of z are marked. If such a node z is found, mark it. Repeat this procedure until there are no more unmarked nodes in the tree all of whose children are marked. The resulting tree is τ_{i+1}.

$$\text{Let } \tau = \bigcup_{i=0}^{\infty} \tau_i.$$

Suppose x is a marked node in τ. If x was added to τ using case 2.a or 2.c, then x occurs within a subtree of τ that is a proof tree. So x must satisfy one of the conditions in Definition 6.49. If x was added to τ *and marked according to case 2.b*, then x is a leaf node in τ and x satisfies condition 2 or 3 of Definition 6.49. Otherwise, x is a non-leaf node in τ, x was added to τ using condition 2.b, and x was marked because all of its children were marked. Looking at the cases used to add the children of x to τ (we add children for each rule for q), x must satisfy condition 2 in Definition 6.49. So if τ is finite and if every node in τ is marked, then τ is a proof tree.

Since cases 2.a–2.c append to node x nodes labeled with a literal from $\mathcal{X}$ or $\mathcal{Y}$ and both of these are finite, it must be the case that the branching factor of τ is finite. So if τ is infinite, then τ must have an infinitely long branch. Consider such a branch. Every node in this branch (other than the top node) must have been added using case 2.b since all the other branches add proof trees which are finite. So every node in the branch must be labeled $-\delta q$ for some literal q. Furthermore, no node in the branch satisfies condition 3 in Definition 6.49 since if it did, it would have been marked when it was added to τ and it would therefore have no children. But since $\mathcal{X}$ is finite, only finitely many literals occur in $\mathcal{X}$. So there must be some literal q such that two different nodes in our infinite branch are labeled $-\delta q$. But then one of these two nodes does satisfy condition 3 of Definition 6.49, which is a contradiction. Therefore, τ is not infinite.

Since τ is not infinite, we can let n be a non-negative integer such that $\tau = \tau_n$. Suppose τ_n has an unmarked node. Since a node must be marked if all its children are marked, τ_n must have an unmarked leaf node x. This node must have been added by case 2.b of our construction, and so we can let q be a literal such that x is labeled $-\delta q$, and $q \in \mathcal{U}_{\mathcal{X}}$. Since x is not marked, it satisfies neither condition 2 or 3 of Definition 6.49. If there is no rule $r \in R$ such that $head(r) = q$, then x satisfies condition 2 of Definition 6.49. So there is a rule $r \in R$ such that $head(r) = q$, and as such one of the cases 2.a-2.c applies to x. So there must be some $m > n$ such x has a child node in τ_m. Then x is not a leaf node in τ_m and x is not a leaf node in τ, a contradiction. Therefore, every node in τ satisfies some condition in Definition 8 and τ is a proof tree.

If α is a limit ordinal, then by Lemma 4.21, there is a $\beta < \alpha$ such that $\mathcal{X}$ is a finite grounded component of $\mathcal{I}_{D,\beta}$. By inductive hypothesis, if $p \in \mathcal{T}_{\mathcal{X}}$, then $D \mid\sim_{MDL} p$, and if $p \in \mathcal{U}_{\mathcal{X}}$, then $D \sim\!\!|_{MDL} p$. □

E.3 Cautious Monotony and Cut for Unprioritized MDL

Cautious Monotony and Cut are shown to hold for MDL for theories without priorities. We prove somewhat more general claims: If $D \mid\!\approx_{MDL} p$, then we may add either $\{\Rightarrow p\}$ or $\{\rightarrow p\}$ to D and not affect the consequences. Note that if we add $s : \{\} \rightarrow p$ to a theory, the conflict sets of D must be altered: if $r \in R_{du}[p]$, $c \in C[r]$, and $c - \{r\} \neq \varnothing$, then we replace c with $c - \{r\}$. If instead we add $s : \{\} \Rightarrow p$ to the theory, then $(c - \{r\}) \cup \{s\}$ is added as an additional conflict set.

Importantly, in order for the below results to obtain, conflict sets cannot be arbitrary. Specifically, it must be the case that if $c \in C[r]$ is a conflict set and $X \in Supp(head(r))$ (*i.e.*, X is a defeasible set of support for $head(r)$), then $c - \{r\} \cup X$ is a conflict set.

Theorem 6.53 (Part I: Cautious Monotony for Unprioritized Theories)**.** *Let $D = \langle R_D, C_D, \varnothing\rangle$ be a defeasible theory. Let $E = \langle R_D \cup \{r_p\}, C_E, \varnothing\rangle$ where $r_p = \{\} \rightarrow p$ or $r_p = \{\} \Rightarrow p$ and C_E is described as above. If $D \mid\!\approx_{MDL} p$ and $D \mid\!\approx_{MDL} q$, then $E \mid\!\approx_{MDL} q$, and if $D \mid\!\approx_{MDL} p$ and $D \approx\!\!|_{MDL} q$, then $E \approx\!\!|_{MDL} q$.*

Proof. Note that if $D \mid\!\approx_{MDL} p$ and $D \mid\!\approx_{MDL} q$ ($D \not\approx_{MDL} q$), there must be a least ordinal α where $p \in \mathcal{T}_{D,\alpha}$ and $q \in \mathcal{T}_{D,\alpha}$ ($q \in \mathcal{U}_{D,\alpha}$). It then suffices to show that for all $\alpha \geq 0$, if $p \in \mathcal{T}_{D,\alpha}$ and $q \in \mathcal{T}_{D,\alpha}$ ($q \in \mathcal{U}_{D,\alpha}$), then $q \in \mathcal{T}_{E,WF}$ ($q \in \mathcal{U}_{E,WF}$). The proof is by induction on the sequence $\mathcal{I}_D$. Suppose that for all $\kappa < \lambda$, if $p \in \mathcal{T}_{D,\kappa}$ and $q \in \mathcal{T}_{D,\kappa}$, then $q \in \mathcal{T}_{E,WF}$, and if $p \in \mathcal{T}_{D,\kappa}$ and $q \in \mathcal{U}_{D,\kappa}$, then $q \in \mathcal{U}_{E,WF}$. We proceed in cases:

(Case 1) Suppose that $q \in \mathcal{T}_{D,\lambda}$ and $p \in \mathcal{T}_{D,\lambda}$ and λ is a successor ordinal. Then there exists a strict or defeasible rule $r \in R_D[q]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$. Since $R_D[q] \subset R_E[q]$, $r \in R_E[q]$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{E,WF}$.

If r is strict, then clearly $q \in \mathcal{T}_{E,WF}$ by definition of T_E and $\mathcal{I}_{E,WF}$.

Suppose r is defeasible. Let $c \in C_D[r]$. There are several cases to consider: (1) $c \in C_E[r]$; (2) $c - \{t\} \in C_E[r]$ (where t is a defeasible rule for p that has been deleted because the rule $r_p = \{\} \rightarrow p$ has been added to E); or (3) $(c - \{t\}) \cup \{r_p\} \in C_E[r]$ (where $r_p = \{\} \Rightarrow p$. We consider each in turn.

(Case 1.1) $c \in C_E[r]$. Since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{r\}$ such that $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. By the inductive hypothesis, $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{E,WF}) \neq \varnothing$.

(Case 1.2) $c - \{t\} \in C_E[r]$. Since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{r\}$ such that $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. If $w \neq t$, then by the inductive hypothesis, $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Suppose $w = t$, then $c - \{w\}$ is a conflict set of $C_E[r]$. Since $D \mid\!\approx_{MDL} p$, there exists a defeasible set X of support for p such that for all $v \in X$, $(\{head(v)\} \cup body(v)) \subseteq \mathcal{T}_{D,\kappa}$ for some $\kappa \geq 0$. As this is so, by coherence we have for each v, $(\{head(v)\} \cup body(v)) \cap \mathcal{U}_{D,\alpha} = \varnothing$ for all $\alpha \geq 0$. Note that because of this, for each $v \in X$, $v \neq w$. However, $(c - \{w\}) \cup X$ is a conflict set of D containing r and so there must be a $w' \in (c - \{r, w\})$ such that $(\{head(w')\} \cup body(w')) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. As such, by the inductive hypothesis, $(\{head(w')\} \cup body(w')) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

(Case 1.3) $(c - \{t\}) \cup \{r_p\} \in C_E[r]$. Since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{r\}$ such that $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. If $w \neq t$, then by the inductive hypothesis, $(head(w) \cup body(w)) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Suppose $w = t$, then $c - \{t\} \cup \{r_p\}$ is a conflict set of $C_E[r]$. Since $D \not\approx_{MDL} p$, then there exists a defeasible set X of support for p such that for all $u \in X$, $(\{head(v)\} \cup body(v)) \subseteq \mathcal{U}_{D,\kappa}$ for some $\kappa \geq 0$. As this is so, for each v, $(\{head(v)\} \cup body(v)) \cap \mathcal{U}_{D,\alpha} = \varnothing$ for all $\alpha \geq 0$. However, $(c - \{w\}) \cup X$ is a conflict set of D containing r, and so there must be a $w' \in (c - \{w, r\})$ such that $(\{head(w')\} \cup body(w')) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. As such, by the inductive hypothesis, $(\{head(w')\} \cup body(w')) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Generalizing on c, there thus exists an $r \in R_{E,sd}[q]$ such that $body(r) \subseteq \mathcal{T}_{E,WF}$ and if $r \in R_{E,d}[q]$, then for all $c \in C_E[r]$, there exists a $s \in c - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{E,WF} \neq \varnothing$. By definition of T_E and $\mathcal{I}_{E,WF}$, $q \in \mathcal{T}_{E,WF}$.

If λ is a limit ordinal and $p \in \mathcal{T}_{D,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{T}_{D,\eta}$. By the inductive hypothesis, $p \in \mathcal{T}_{E,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{D,\lambda}$ and $p \in \mathcal{T}_{D,\lambda}$ and λ is a successor ordinal, and let a be any literal such that $a \in \mathcal{U}_{D,\lambda}$. Since $\mathcal{I}_\lambda$ is coherent and $D \not\approx_{MDL} p$, $a \neq p$ and so $R_D[a] = R_E[a]$. Since $a \in \mathcal{U}_{D,\lambda}$, if $r \in R_D[a]$ and r is strict, then $body(r) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$. Since U_D is monotonic, it follows that $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$.

Suppose r is defeasible. Then $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a $c \in C_D[r]$ such that for each $w \in c - \{r\}$, $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{D,\lambda-1}$. By the inductive hypothesis, for each w, $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{E,WF}$. Since $c \in C_D[r]$, then either (1) $c' = c \in C_E$, or (2) $c' = (c - \{t\}) \in C_E$, or (3) $c' = (c - \{t\}) \cup \{r_p\} \in C_E$. Recall that $body(r_p) = \varnothing$ and that (by case 1 above) since $p \in \mathcal{T}_{D,\lambda}$, we have $p \in \mathcal{T}_{E,WF}$. Then regardless of whether 1,2, or 3 holds, we have for all $v \in c' - \{r\}$, $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{E,WF}$. Since $R_D[a] = R_E[a]$, generalizing on r we have that for each strict or defeasible $r \in R_E[a]$, either $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else r is defeasible and there exists a conflict set $c \in C_E[r]$ such that for each $w \in c - \{r\}$ $\{head(w)\} \cup body(w) \subseteq \mathcal{T}_{E,WF}$. Generalizing on a, it can be seen that $\mathcal{U}_{D,\lambda}$ is unfounded *wrt* E and $\mathcal{I}_{E,WF}$. By definition of U_E and $\mathcal{I}_{E,WF}$, we have $q \in \mathcal{U}_{E,WF}$.

If λ is a limit ordinal and $p \in \mathcal{U}_{D,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{U}_{D,\eta}$. By the inductive hypothesis, $p \in \mathcal{U}_{E,WF}$. □

Theorem 6.53 (Part II: Cut for Unprioritized Theories)**.** *Let L be MDL, D an unprioritized defeasible theory such that $D \mid\!\approx_{MDL} p$. Let $E = \langle R_D \cup X, C_D, \varnothing\rangle$ where $X = \{\rightarrow p\}$ or $X = \{\Rightarrow p\}$. If $E \mid\!\approx_{MDL} q$, then $D \mid\!\approx_{MDL} q$, and if $E \not\approx_{MDL} q$, then $D \not\approx_{MDL} q$.*

Proof. Note that if $D \mid\!\approx_{MDL} p$, then by cautious monotony, we have $E \mid\!\approx_{MDL} p$. Furthermore, if $E \mid\!\approx_{MDL} p$ and $E \mid\!\approx_{MDL} q$ ($E \not\approx_{MDL} q$), there must be a least ordinal κ where $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\kappa}$ ($q \in \mathcal{U}_{E,\kappa}$). It thus suffices to show that for all $\kappa \geq 0$, if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\alpha}$ ($q \in \mathcal{U}_{E,\alpha}$), then $q \in \mathcal{T}_{D,WF}$ ($q \in \mathcal{U}_{D,WF}$). Suppose that for all $\kappa < \lambda$, if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\kappa}$, then $q \in \mathcal{T}_{D,WF}$, and if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{U}_{E,\kappa}$, then $q \in \mathcal{U}_{D,WF}$. We proceed in cases:

(Case 1) Suppose $q \in \mathcal{T}_{E,\lambda}$ and $p \in \mathcal{T}_{E,\lambda}$ and that λ is a successor ordinal. If $q = p$, then by assumption $D \mid\!\approx p$. Let $q \neq p$. Then there exists a rule $r \in R_E[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. Since $q \neq p$, $r \in R_D[q]$ and by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$. If r is strict, then clearly $q \in \mathcal{T}_{D,WF}$ by definition of T_D and $\mathcal{I}_{D,WF}$.

Suppose r is defeasible and let $c \in C_E[r]$. Then since $q \in \mathcal{T}_{E,\lambda}$ there exists a $w \in c - \{r\}$ such that $\{head(w)\} \cup body(w) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. By the inductive hypothesis, $\{head(w)\} \cup body(w) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Since $r_p \in R_E$, $p \in \mathcal{T}_{E,\lambda}$ (by assumption) and $body(r_p) = \varnothing$, it must be the case that $w \neq r_p$. Note that every conflict set $c \in C_E$ is constructed from some $c' \in C_D$, and if $s \in c$ and $s \neq r_p$, then $s \in c'$. Generalizing on c, for each $c' \in C_D[r]$, there exists a $w \in c' - \{r\}$ such that $\{head(w)\} \cup body(w) \cap \mathcal{U}_{D,WF} \neq \varnothing$. And so we have $r \in R_{D,d}[q]$ and $body(r) \subseteq \mathcal{T}_{D,WF}$ and for all $c' \in C_D[r]$, there is a $w \in c' - \{r\}$ such that $\{head(w)\} \cup body(w) \cap \mathcal{U}_{D,WF} \neq \varnothing$. By definition of T_D and $\mathcal{I}_{D,WF}$, $q \in \mathcal{T}_{D,WF}$.

If λ is a limit ordinal and $p \in \mathcal{T}_{E,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{T}_{E,\eta}$. By the inductive hypothesis, $p \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{E,\lambda}$ and $p \in \mathcal{T}_{E,\lambda}$. Let a be any literal in $\mathcal{U}_{E,\lambda}$. By coherence $a \neq p$. If $r \in R_E[a]$ and r is strict, then $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. Since U_E is monotonic, it follows that $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$. If r is defeasible, then $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a $c \in C_E[r]$ such that for each $w \in c - \{r\}$ $\{head(w)\} \cup body(w) \subseteq \mathcal{T}_{E,\lambda-1}$. There are three cases to consider: (1) $c \in C_D$; (2) $c \cup \{t\} \in C_D$ where $t \in R_d[p]$; and (3) $(c - \{r_p\}) \cup \{t\} \in C_D$, where

$t \in R_d[p]$. Note that if $X \in Supp(p)$ and $c' \in C_D[s]$ where $head(s) = p$, then $c' - \{s\} \cup X$ is a conflict set of D. Since $D \mid\!\approx_{MDL} p$ there exists a $X \in Supp(p)$ such that for each rule $y \in X$, $\{head(y)\} \cup body(y) \subseteq \mathcal{T}_{D,WF}$. Thus in cases 2 and 3 we have for all rules $v \in (c - \{r, t\}) \cup X)$, $\{head(v)\} \cup body(v) \subseteq \mathcal{T}_{D,WF}$.

Since $a \neq p$ we have $R_E[a] = R_D[a]$. Generalizing on r we have for each strict or defeasible $r \in R_D[a]$ either $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$ or else r is defeasible and there exists a conflict set $c \in C_D[r]$ such that for each $w \in c - \{r\}$ $\{head(w)\} \cup body(s) \subseteq \mathcal{T}_{D,WF}$. Generalizing on a, it can be seen that $\mathcal{U}_{E,\lambda}$ is an unfounded set *wrt* D and $\mathcal{I}_{D,WF}$. Thus $q \in \mathcal{U}_{D,WF}$.

If λ is a limit ordinal and $p \in \mathcal{U}_{E,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{U}_{E,\eta}$. By the inductive hypothesis, $p \in \mathcal{U}_{D,WF}$. □

E.4 Rule Simplification

Theorem 6.53 (Rule Simplification)**.** *Let D a defeasible theory such that $D \mid\!\approx_{MDL} p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) $D \mid\!\approx_{MDL} q$ iff $E \mid\!\approx_{MDL} q$.
(2) $D \approx\!\!\mid_{MDL} q$ iff $E \approx\!\!\mid_{MDL} q$.

The above theorem follows directly from the two Lemmas proven below.

Lemma E.5. *Let D a defeasible theory such that $D \mid\!\approx_{MDL} p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) If $D \mid\!\approx_{MDL} q$ then $E \mid\!\approx_{MDL} q$.
(2) If $D \approx\!\!\mid_{MDL} q$ then $E \approx\!\!\mid_{MDL} q$.

Proof. Note that if $D \mid\!\approx_{MDL} p$ and $D \mid\!\approx_{MDL} q$ ($D \approx\!\!\mid_{MDL} q$), there must be a least ordinal α where $p \in \mathcal{T}_{D,\alpha}$ and $q \in \mathcal{T}_{D,\alpha}$ ($q \in \mathcal{U}_{D,\alpha}$). As in the other proofs, we will show that for all $\alpha \geq 0$, if $p \in \mathcal{T}_{D,\alpha}$ and $q \in \mathcal{T}_{D,\alpha}$ ($q \in \mathcal{U}_{D,\alpha}$), then $q \in \mathcal{T}_{E,WF}$ ($q \in \mathcal{U}_{E,WF}$). The proof is by induction

on the sequence $\mathcal{I}_D$. Suppose that for all $\kappa < \lambda$, if $p \in \mathcal{T}_{D,\kappa}$ and $q \in \mathcal{T}_{D,\kappa}$, then $q \in \mathcal{T}_{E,WF}$, and if $p \in \mathcal{T}_{D,\kappa}$ and $q \in \mathcal{U}_{D,\kappa}$, then $q \in \mathcal{U}_{E,WF}$. The claim obviously holds for $\kappa = 0$. We proceed in cases:

(Case 1) Suppose $q \in \mathcal{T}_{D,\lambda}$ and $p \in \mathcal{T}_{D,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $r \in R_{D,sd}[q]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$. If $r \neq t$, then $r \in R_E$ and by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{E,WF}$. If $r = t$, then since $body(t') \subseteq body(t)$, $body(t') \subseteq \mathcal{T}_{E,WF}$.

Suppose r is strict. If $r \neq t$ then $r \in R_E$, and since $body(r) \subseteq \mathcal{T}_{E,WF}$, $q \in \mathcal{T}_{E,WF}$ by definition of T_E and $\mathcal{I}_{E,WF}$. If r is strict and $r = t$, then t' is strict; since $body(t') \subseteq \mathcal{T}_{E,WF}$, clearly $q \in \mathcal{T}_{E,WF}$.

Suppose r is defeasible and $r = t$. Let $c \in C_D[t]$. Since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{t\}$ such that either $w \prec t$ or $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$. The counterpart to c in C_E is $c' = c - \{t\} \cup \{t'\}$ (and since t is defeasible, so is t'). As such, for all $w \in c - \{t'\}$, $w \in R_E$ and if $w \prec t$, then $w \prec t'$.

Suppose r is defeasible and $r \neq t$. Let $c \in C_D[r]$ and c' be its counterpart in C_E. Since $q \in \mathcal{T}_{D,\lambda}$ there exists a $w \in c - \{r\}$ such that either $w \prec r$ or $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$. If $w \neq t$, then $w \in R_E$. If $w = t$, then since $p \in \mathcal{T}_{D,WF}$, if $(\{head(t)\} \cup body(t)) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$, there must exist a $v \in (\{head(t)\} \cup body(t))$ such that $v \neq p$ and $v \in \mathcal{U}_{D,\kappa-1}$. Since $body(t') = body(t) - \{p\}$, then if $(\{head(t)\} \cup body(t)) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$, then $(\{head(t')\} \cup body(t')) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$. Thus, regardless of whether $w = t$ or not, there exists a rule $s \in c' - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$ or $w \prec r$.

Note that there is a 1–1 correspondence between C_D and C_E. Generalizing on c (and hence c'), if $r \in R_E$, then for all $c' \in C_E[r]$, there is a $w \in c' - \{r\}$ such that $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,\kappa-1} \neq \varnothing$ or $w \prec r$. By the inductive hypothesis we have and $body(r) \subseteq \mathcal{T}_{E,WF}$, and for all $c' \in C_E[r]$, there is a $w \in c' - \{r\}$ such that $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or $w \prec r$. By definition of T_E and $\mathcal{I}_{E,WF}$, $p \in \mathcal{T}_{E,WF}$. If $r \notin R_E$, then $r = t$, and since $body(r) \subseteq \mathcal{T}_{D,\kappa-1}$ and $body(t') \subset body(t)$, by inductive hypothesis $body(t) \subseteq \mathcal{T}_{E,WF}$. Furthermore, for all $c' \in C_E[t']$,

there is a $w \in c' - \{t'\}$ such that $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or $w \prec t'$. By definition of T_E and $\mathcal{I}_{E,WF}$, $p \in \mathcal{T}_{E,WF}$.

If λ is a limit ordinal and $p \in \mathcal{T}_{D,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{T}_{D,\eta}$. By the inductive hypothesis, $p \in \mathcal{T}_{E,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{D,\lambda}$ and $p \in \mathcal{T}_{D,\lambda}$, and let $a \in \mathcal{U}_{D,\lambda}$. Suppose that λ is a successor ordinal. Since $\mathcal{I}_\lambda$ is coherent, $a \neq p$ and so $R_D[a] = R_E[a]$. Since $a \in \mathcal{U}_{D,\lambda}$, if $r \in R_s[a]$, then $body(r) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$. Since U_D is monotonic, it follows that $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$. If $r \in R_d[a]$, then $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a $c \in C[r]$ such that for each $w \in c - \{r\}$ $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{D,\lambda-1}$ and $w \not\prec r$. If $t \notin c - \{r\}$ then $c' = c$, and every rule in $c' - \{r\}$ is a rule of E, and so for each $w \in c' - \{r\}$ (by the inductive hypothesis) $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{E,WF}$ and $w \not\prec r$. If $t \in c - \{r\}$, then since $body(t') \subset body(t)$ and $t \not\prec r$ *iff* $t' \not\prec r$, we have $(\{head(t')\} \cup body(t')) \subseteq \mathcal{T}_{D,\lambda-1}$ and $t' \not\prec r$. And so (as before) for each $w \in c' - \{r\}$ (by the inductive hypothesis) $\{head(w)\} \cup body(w) \subseteq \mathcal{T}_{E,WF}$ and $w \not\prec r$.

Generalizing on r, for each $r \in R_{E,sd}[a]$, either $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else r is defeasible and there exists a conflict set $c \in C_E[r]$ such that for each $w \in c - \{r\}$, $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{E,WF}$ and $w \not\prec r$. Generalizing on a, it can be seen that $\mathcal{U}_{D,\lambda}$ is unfounded *wrt* E and $\mathcal{I}_{E,WF}$. Thus $q \in \mathcal{U}_{E,WF}$.

If λ is a limit ordinal and $p \in \mathcal{U}_{D,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{U}_{D,\eta}$. By the inductive hypothesis, $p \in \mathcal{U}_{E,WF}$. □

Lemma E.6. *Let D a defeasible theory such that $D \mid\!\approx_{MDL} p$. Let t be any rule such that $p \in body(t)$, t' the rule obtained by deleting p from $body(t)$, and let E be the theory obtained by replacing t with t'. For all $q \in Lit_D$,*

(1) If $E \mid\!\approx_{MDL} q$ then $D \mid\!\approx_{MDL} q$.
(2) If $E \approx\!\!|_{MDL} q$ then $D \approx\!\!|_{MDL} q$.

Proof. Note that if $D \mid\!\approx_{MDL} p$, then by Lemma E.5, we have $E \mid\!\approx_{MDL} p$. Furthermore, if $E \mid\!\approx_{MDL} p$ and $E \mid\!\approx_{MDL} q$ ($E \approx\!\!|_{MDL} q$), there must be a least ordinal κ where $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\kappa}$ ($q \in \mathcal{U}_{E,\kappa}$). It thus suffices to show that for all $\kappa \geq 0$, if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\alpha}$

($q \in \mathcal{U}_{E,\alpha}$), then $q \in \mathcal{T}_{D,WF}$ ($q \in \mathcal{U}_{D,WF}$). Suppose that for all $\kappa < \lambda$, if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{T}_{E,\kappa}$, then $q \in \mathcal{T}_{D,WF}$, and if $p \in \mathcal{T}_{E,\kappa}$ and $q \in \mathcal{U}_{E,\kappa}$, then $q \in \mathcal{U}_{D,WF}$. We proceed in cases:

(Case 1) Suppose $q \in \mathcal{T}_{E,\lambda}$ and $p \in \mathcal{T}_{E,\lambda}$ and λ is a successor ordinal. If $q = p$, then by assumption $D \mid\!\approx_{MDL} p$. Suppose $q \neq p$. Then there exists a rule $r \in R_E[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. Since $q \neq p$, $r \in R_D[q]$ and by the inductive hypothesis $body(r) \subseteq \mathcal{T}_{D,WF}$.

If r is strict, then clearly $q \in \mathcal{T}_{D,WF}$ by definition of T_D and $\mathcal{I}_{D,WF}$. Suppose r is defeasible. Let $c \in C_E[r]$. Then since $q \in \mathcal{T}_{E,\lambda}$ there exists a $w \in c - \{r\}$ such that either $w \prec r$ or $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{E,\kappa-1} \neq \varnothing$. By the inductive hypothesis, if $\{head(w)\} \cup body(w) \cap \mathcal{U}_{E,\kappa-1} \neq \varnothing$, then $\{head(w)\} \cup body(w) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Let $c' \in C_D[r]$ be the conflict set corresponding to c in D. If $c = c'$, then we need examine it no further. If $c \neq c'$, then $t' \in c$ and $t \in c'$. If $w = t'$, then either $t' \not\prec r$ or else $(\{head(t')\} \cup body(t')) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Since $body(t') \subset body(t)$ and $head(t') = head(t)$, we have either $t \not\prec r$ or $(\{head(t)\} \cup body(t)) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Thus we are certain that there exists a $w \in c' - \{r\}$ such that $w \not\prec r$ or $(\{head(w)\} \cup body(w)) \cap \mathcal{U}_{D,WF} \neq \varnothing$.

Generalizing on c, since there is a 1–1 correspondence between the conflict sets of E and D, we have $r \in R_D$, $body(r) \subseteq \mathcal{T}_{D,WF}$ and for each $c' \in C_D[r]$, there is a $w \in c' - \{r\}$ such that $\{head(w)\} \cup body(w) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or $w \prec r$. By definition of T_D and $\mathcal{I}_{D,WF}$, $q \in \mathcal{T}_{D,WF}$.

If λ is a limit ordinal and $p \in \mathcal{T}_{E,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{T}_{E,\eta}$. By the inductive hypothesis, $p \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{E,\lambda}$ and $p \in \mathcal{T}_{E,\lambda}$ and let a be any literal such that $a \in \mathcal{U}_{E,\lambda}$. Suppose that λ is a successor ordinal.

Since we have $p \in \mathcal{T}_{E,\lambda}$, by coherence we may assume that $a \neq p$. Thus $R_D[a] = R_E[a]$. If $r \in R_s[a]$, then $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. Since U_E is monotonic, it follows that $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$. If $r \in R_d[a]$, then $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a $c \in C_E[r]$ such that for each $w \in c - \{r\}$, $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{E,\lambda-1}$ and $w \not\prec r$. By inductive hypothesis, $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{D,WF}$. Observe that if $w \neq t'$, then $w \in R_D$. If $w = t'$, since $p \in \mathcal{T}_{D,WF}$ and $body(t') \subset body(t)$ and $head(t') = head(t)$, we have $(\{head(t)\} \cup body(t)) \subseteq \mathcal{T}_{D,WF}$

and $t \not\prec r$. Thus, regardless of whether $t' \in c$ or not, we are assured that in the counterpart conflict set $c' \in C_D$, for each $s \in (c' - \{r\})$, $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$.

Generalizing on r, for each $r \in R_{D,sd}[a]$, either $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else r is defeasible and there exists a conflict set $c \in C_D[r]$ such that for each $w \in c - \{r\}$ $(\{head(w)\} \cup body(w)) \subseteq \mathcal{T}_{D,WF}$ and $w \not\prec r$. Generalizing on a, it can be seen that $\mathcal{U}_{E,\lambda}$ is unfounded *wrt* D and $\mathcal{I}_{D,WF}$. Thus $q \in \mathcal{U}_{D,WF}$.

If λ is a limit ordinal and $p \in \mathcal{U}_{E,\lambda}$, then there is a least ordinal $\eta < \lambda$ such that $p \in \mathcal{U}_{E,\eta}$. By the inductive hypothesis, $p \in \mathcal{U}_{D,WF}$. □

E.5 Rule Elimination

In this section we show that we may delete the rules of unfounded literals without altering the consequences of the theory. If we delete a rule t, then we must also delete all conflict sets in which t appears.

Theorem 6.54 (Rule Elimination)**.** *Let $D = \langle R_D, C_D, \prec_D \rangle$ be a prioritized defeasible theory such that $D \approx\!|_{MDL}\ p$. Let t be any rule such that $p \in body(t)$, and let $E = \langle R_D - \{t\}, C_E, \prec_E \rangle$, where $C_E = C_D - \{c | t \in c\}$ and $\prec_E = \prec_D - \{a \prec b | a \prec b \wedge t = a \vee t = b\}$. For all $q \in Lit_D$,*

(1) $D \not\approx_{MDL} q$ iff $E \not\approx_{MDL} q$.
(2) $D \approx\!|_{MDL} q$ iff $E \approx\!|_{MDL} q$.

Proof. Follows directly from the below two lemmas. □

Lemma E.7. *Let $D = \langle R_D, C_D, \prec_D \rangle$ a prioritized defeasible theory such that $D \approx\!|_{MDL}\ p$. Let t be any rule such that $p \in (\{head(t)\} \cup body(t))$, and let $E = \langle R_D - \{t\}, C_E, \prec_E \rangle$, where $C_E = C_D - \{c | t \in c\}$ and $\prec_E = \prec_D - \{a \prec b | a \prec b \wedge t = a \vee t = b\}$. For all $q \in Lit_D$,*

(1) If $D \not\approx_{MDL} q$ then $E \not\approx_{MDL} q$.
(2) If $D \approx\!|_{MDL} q$ then $E \approx\!|_{MDL} q$.

Proof. The proof is by induction on the sequence $\mathcal{I}_{D,0}, \mathcal{I}_{D,1}, \ldots$. The hypothesis is trivially satisfied for $\mathcal{T}_{D,0}$. Suppose for all $\kappa < \lambda$, if $q \in \mathcal{T}_{D,\kappa}$ then $q \in \mathcal{T}_{E,WF}$, and if $q \in \mathcal{U}_{D,\kappa}$ then $p \in \mathcal{U}_{E,WF}$. We will treat the cases where $q \in \mathcal{T}_{D,\lambda}$ and $q \in \mathcal{U}_{D,\lambda}$ separately.

(Case 1) Suppose $q \in \mathcal{T}_{D,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $r \in R_D[q]$ such that $body(r) \subseteq \mathcal{T}_{D,\lambda-1}$. By monotonicity of T_D, $body(r) \subseteq \mathcal{T}_{D,\lambda}$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{E,WF}$. Since $(\{head(r)\} \cup body(r)) \subseteq \mathcal{T}_{D,\lambda}$, $r \neq t$, and so $r \in R_E$. If r is strict, then clearly $q \in \mathcal{T}_{E,WF}$ by definition of T_E and $\mathcal{T}$. Suppose r is defeasible. Let $c \in C[r]$. Then since $q \in \mathcal{T}_{D,\lambda}$ there exists a $s \in c - \{r\}$ such that either $s \prec r$ or $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{D,\lambda-1} \neq \varnothing$. If $s = t$, Then $c \notin C_E[r]$. If $s \neq t$, then $s \prec r$ or by the inductive hypothesis $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{E,WF} \neq \varnothing$.

Generalizing on c, we have $r \in R_E$ and $body(r) \subseteq \mathcal{T}_{E,WF}$ and for each conflict set $c \in C_E[r]$, there exists a $s \in c - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{E,WF} \neq \varnothing$ or else $s \prec r$. By definition of and T_E and $\mathcal{T}_{E,WF}$, $q \in \mathcal{T}_{E,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\eta < \lambda$ such that $q \in \mathcal{T}_{D,\eta}$ and so by the inductive hypothesis $q \in \mathcal{T}_{E,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{D,\lambda}$, λ is a successor ordinal, and let a be any literal in $\mathcal{U}_{D,\lambda}$. Suppose $r \in R_{D,s}[a]$, Since $a \in \mathcal{U}_{D,\lambda}$, then $body(r) \cap (\mathcal{U}_{D,\lambda} \cup \mathcal{U}_{D,\lambda-1}) \neq \varnothing$. Since U_D is monotonic, it follows that $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$.

Suppose $r \in R_{D,d}[a]$, then $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a $c \in C[r]$ and for each $s \in c - \{r\}$ $\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{D,\lambda-1}$ and $s \not\prec r$. For any such s, since $\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{D,\lambda-1}$, it cannot be the case that $s = t$. Thus for each $s \in c - \{r\}$, $s \in R_E[w]$ and $\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{D,\lambda-1}$ and $s \not\prec r$.

Note that $R_E[a] \subseteq R_D[a]$ and for any r, $C_E[r] \subseteq C_D[r]$. Generalizing on r, for each $r \in R_{E,s}[a]$, $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$. For each $r \in R_{E,d}[a]$, either $body(r) \cap \mathcal{U}_{D,\lambda} \neq \varnothing$, or else there exists a conflict set $c \in C_E[r]$ such that for each $s \in c - \{r\}$, $s \in R_E[w]$ and $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{E,WF}$ and $s \not\prec r$. Generalizing on a,we see that $\mathcal{U}_{D,\lambda}$ is unfounded *wrt* E and $\mathcal{I}_{E,WF}$. It follows by definition of U_E and $\mathcal{I}_{E,WF}$, that $q \in \mathcal{U}_{E,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\eta < \lambda$ such that $q \in \mathcal{U}_{D,\eta}$ and so by the inductive hypothesis $q \in \mathcal{U}_{E,WF}$. □

Lemma E.8. *Let $D = \langle R_D, C_D, \prec_D \rangle$ be a prioritized defeasible theory such that $D \not\approx\!|_{MDL}\ p$. Let t be any rule such that $p \in \{head(t) \cup body(t))$, and let $E = \langle R_D - \{t\}, C_E, \prec_E \rangle$, where $C_E = C_D - \{c | t \in c\}$ and $\prec_E = \prec_D - \{a \prec b | a \prec b \wedge t = a \vee t = b\}$. For all $q \in Lit_D$,*

(1) If $E \not\approx_{MDL} q$ then $D \not\approx_{MDL} q$.
(2) If $E \not\approx\!|_{MDL} q$ then $D \not\approx\!|_{MDL} q$.

Proof. The proof is by induction on the sequence $\mathcal{I}_{E,0}, \mathcal{I}_{E,1}, \ldots$. The hypothesis is trivially satisfied for $\mathcal{T}_{E,0}$. Suppose for all $\kappa < \lambda$, if $q \in \mathcal{T}_{E,\kappa}$ then $q \in \mathcal{T}_{D,WF}$, and if $q \in \mathcal{U}_{E,\kappa}$ then $p \in \mathcal{U}_{D,WF}$. We will treat the cases where $q \in \mathcal{T}_{E,\lambda}$ and $q \in \mathcal{U}_{E,\lambda}$ separately.

(Case 1) Suppose $q \in \mathcal{T}_{E,\lambda}$ and that λ is a successor ordinal. Then there exists a rule $r \in R_E[q]$ such that $body(r) \subseteq \mathcal{T}_{E,\lambda-1}$. By the inductive hypothesis, $body(r) \subseteq \mathcal{T}_{D,WF}$. Since $r \in R_E$ and $R_E \subseteq R_D$ it must be that $r \neq t$ and $r \in R_D$. If r is strict, then clearly $q \in \mathcal{T}_{D,WF}$ by definition of T_D and $\mathcal{T}_{D,WF}$.

Suppose r is defeasible. Let $c \in C_E[r]$. Then since $q \in \mathcal{T}_{E,\lambda}$ there exists a $s \in c - \{r\}$ such that either $s \prec r$ or $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{E,\lambda-1} \neq \varnothing$. If the latter, then by the inductive hypothesis, $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{D,WF} \neq \varnothing$.

Note that $C_E[r] \subseteq C_D[r]$, and for each $c' \in C_D[r]$ such that $c' \notin C_E[r]$, we have $t \in c'$. However, it is already known that $(\{head(t)\} \cup body(t)) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Generalizing on c, we have $r \in R_D[q]$, $body(r) \subseteq \mathcal{T}_{D,WF}$, and for each conflict set $c \in C_D[r]$, there exists a $s \in c - \{r\}$ such that $(\{head(s)\} \cup body(s)) \cap \mathcal{U}_{D,WF} \neq \varnothing$ or else $s \prec r$. By definition of and T_D and $\mathcal{T}_{D,WF}$, $q \in \mathcal{T}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\eta < \lambda$ such that $q \in \mathcal{T}_{E,\eta}$ and so by the inductive hypothesis $q \in \mathcal{T}_{D,WF}$.

(Case 2) Now suppose that $q \in \mathcal{U}_{E,\lambda}$, λ is a successor ordinal, and let a be any literal in $\mathcal{U}_{E,\lambda}$. Suppose $r \in R_{E,s}[a]$. Since $a \in \mathcal{U}_{E,\lambda}$, then $body(r) \cap (\mathcal{U}_{E,\lambda} \cup \mathcal{U}_{E,\lambda-1}) \neq \varnothing$. Since U_E is monotonic, it follows that $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$.

Suppose $r \in R_{E,d}[a]$, then $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a $c \in C[r]$ such that for each $s \in c - \{r\}$ we have $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{E,\lambda-1}$ and $s \not\prec r$. Since $R_E \subseteq R_D$, each such s is

a member of R_D. By the inductive hypothesis, for each such s, $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$.

Note that if $t \in R_D[a]$, then $R_E[a] = R_D[a] - \{t\}$, and it is already known that $(\{head(t)\} \cup body(t)) \cap \mathcal{U}_{D,WF} \neq \varnothing$. Generalizing on r, we may then infer that for each $r \in R_{D,s}[a]$, $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$. For each $r \in R_{D,d}[a]$, either $body(r) \cap \mathcal{U}_{E,\lambda} \neq \varnothing$, or else there exists a conflict set $c \in C[r]$ such that for each $s \in c - \{r\}$, $(\{head(s)\} \cup body(s)) \subseteq \mathcal{T}_{D,WF}$ and $s \not\prec r$. Generalizing on a,we see that $\mathcal{U}_{E,\lambda}$ is unfounded *wrt* D and $\mathcal{I}_{D,WF}$. It follows by definition of U_D and $\mathcal{I}_{D,WF}$, that $q \in \mathcal{U}_{D,WF}$.

If λ is not a successor ordinal, then there is a least successor ordinal $\eta < \lambda$ such that $q \in \mathcal{U}_{E,\eta}$ and so by the inductive hypothesis $q \in \mathcal{U}_{D,WF}$. □