Defeasible Logics

P. Geerts, E. Laenens
Dept. of Mathematics and Computer Science
University of Antwerp

D. Vermeir
Dept. of Computer Science
Free University of Brussels, VUB

Contents

0.1 Introduction 1
0.2 Pearl's system Z 4
0.2.1 Deriving a natural ordering of defaults 5
0.2.2 System Z^+: resolving remaining ambiguities by explicit means 7
0.3 Conditional entailment 8
0.4 The argument-based system of Simari and Loui 11
0.5 Brewka's preferred subtheories 16
0.6 Ordered logic and basic defeasible logic 19
0.6.1 Implicit version of Nute's basic defeasible logic 20
0.6.2 Ordered logic 24
0.6.3 Explicit version of Nute's basic defeasible logic 29
0.6.4 The preemption problem 31
0.6.5 Ryan's formalism of ordered theories presentations 32
0.7 Summary and discussion 33
Bibliography 34

0.1 Introduction

In this chapter, we will concentrate on the defeasible approach to nonmonotonic reasoning, as opposed to the minimalist approach and the fixpoint approach. Formalisms following the minimalist approach, like circumscription [McC80, McC86, Lif85], look at models of a classical theory that are minimal with respect to some set of predicates occurring in the theory. Fixpoint formalisms include McDermott's and Doyle's nonmonotonic logic [MD80], Reiter's default logic [Rei80], and Moore's autoepistemic logic [Moo84, Moo88]. Default rules in these systems involve a special condition for application, which is explained in the proof theory. By applying rules in an arbitrary order, some other rules may become blocked, and eventually, a fixpoint is derived from where no new conclusions can be reached. At a fixpoint, every default is either inapplicable or applied. The same holds for minimal models.

In the defeasible approach, defaults are treated quite differently. The intent of a default $A \to p$ is that p will normally be derivable from a theory containing this default whenever A is derivable. However, it is possible to have a theory containing $A \to p$ from which both A and $\neg p$ can be derived. If this is the case, the rule $A \to p$ is said to be *defeated*. Rules which can be defeated are called *defeasible rules* and logics using defeasible rules are called *defeasible logics*. In a defeasible logic formalism, extensions of theories are formed, in which each rule is either inapplicable, applied or defeated.

In case of a conflict, defeasible logics usually rely on some kind of ordering on defeasible rules or sets of rules in an attempt to resolve this conflict. This prioritization can be implicitly present in the knowledge base, in which case it is based on some notion of specificity, or explicitly given by the user. Both approaches have their pros and cons, which will be discussed at the end of this chapter.

Whenever the ordering on rules does not allow to resolve the conflict, a defeasible logic formalism can adopt either a *credulous* or a *skeptical* strategy. Using the credulous strategy, an arbitrary rule among the competing but incomparable ones is chosen for application. A credulous reasoner wants to draw as much conclusions as possible, so that he prefers to apply one of these rules instead of concluding nothing. As a result, this credulous strategy leads to multiple extensions. Following the skeptical strategy, no conclusion is drawn, and the conflicting rules are said to defeat each other. A skeptical reasoner wants to draw a conclusion only when he's very sure about it, so that in case of doubt, he doesn't conclude anything. This strategy always yields a unique extension. Another way to arrive at a unique extension is to be credulous, derive multiple extensions and then take the intersection of these credulous extensions. In this case, we obtain the conclusions of which we can be very sure, because they are true in every possible world. This approach can be considered as being skeptical after taking into account all possibilities, and is therefore called the *indirectly skeptical* approach. The indirectly skeptical approach usually seems to fit closer to our intuition than the directly skeptical approach, but it is also the most costly regarding computations.

Formalisms following a directly skeptical approach can differ in the way they deal with ambiguities. Whenever there is an ambiguity about a proposition p which cannot be solved by considering priorities, p will not be a conclusion because of the skeptical attitude. However, a formalism can forget about this ambiguity, and therefore also forget that there was a reason to believe p, or it can register p as an ambiguous proposition which can still interfere with other conclusions. The first approach will be called *ambiguity-blocking*, while the second one is *ambiguity-propagating* [Ste92].

In order to treat all formalisms alike, we will make some notational conventions which we will adhere to throughout this chapter. A literal is a propositional constant p or the negation $\sim p$ of a propositional constant; p and $\sim p$ are complements of each other. Where p is any literal, we denote the complement of p as $\neg p$. Where A is a finite set of literals and p is a literal, a *strict rule* is denoted $A \Rightarrow p$, and a *defeasible rule* is denoted $A \rightarrow p$. Strict rules are sometimes referred to as sentences or necessary facts, while defeasible rules are also called default rules or defaults. Some formalisms allow a third kind of rule: a *defeater*, denoted $A \leadsto p$. A defeater can defeat a defeasible rule, but can never support inferences directly. A defeater will therefore also be called an *interfering rule*. An example of a strict rule is "Everything in Brussels is in Belgium"'. "Birds fly" is a defeasible rule, because it can be defeated when we acquire more information. An example of an interfering rule is “A sick bird might not fly”. We usually omit the set brackets when the antecedent set has only one member, and we usually omit an empty antecedent set altogether. Thus $\{p\} \rightarrow q$ is usually written $p \rightarrow q$ and $\emptyset \rightarrow q$ is usually written $\rightarrow q$. Antecedents of strict rules and defeaters are non-empty sets, but defeasible rules might have empty antecedents. A defeasible rule with an empty antecedent can be considered as a *presumption*. For a rule r, we use the notations $B(r)$ for the *body* of the rule and $H(r)$ for the *head*.

Because some formalisms consider the body of a rule as a conjunction of literals instead of a set, we also introduce the notation $CB(r)$ for the *conjuncted body* of r, i.e. the conjunction of literals present in the body of r. Put otherwise, when r is a rule $A \rightarrow p$, then $B(r) = A$, $H(r) = p$ and $CB(r) = \bigwedge_{a \in A} a$. Whenever a formalism is defined for a first-order language, we will simply consider the grounded instances of the rules containing variables. Although some formalisms also allow more complex rules involving e.g. disjunctions, we will restrict the rules in this discussion to the ones described above.

A *knowledge base* is a set of rules R. Some defeasible logics consider all rules to be defeasible, others also allow strict rules and occasionally, interfering rules can be found. When necessary, we will refer to the strict, defeasible and interfering parts of R as R_s, R_d and R_i, i.e.

$$R_s = \{A \Rightarrow p \mid A \Rightarrow p \in R\}$$

$$R_d = \{A \rightarrow p \mid A \rightarrow p \in R\}$$

and

$$R_i = \{A \rightsquigarrow p \mid A \rightsquigarrow p \in R\}$$

A *defeasible theory* T is a knowledge base R together with a (possibly empty) set O containing literals, representing the *observations*. Observations are also called evidences or contingent facts.

Sometimes, some kind of additional structure is added to represent explicit priorities. In this case, the knowledge bases and defeasible theories are said to be *ordered* (or *prioritized*). From a technical point of view, ordered defeasible theories can live without strict rules and observations: their impact can be simulated by giving the corresponding defeasible rules top priority. Although the additional structure for representing explicit priorities can take different forms, we usually can rewrite a defeasible theory with ordering as a general ordered theory $(\Omega, \leq, R, f)$, where $(\Omega, \leq)$ is a finite totally or partially ordered set of nodes, R a finite set of defeasible rules, and f a function assigning a set of rules to each node. In this framework, strict rules and observations can be treated the same way as defeasible rules, provided that they are assigned to a node with top priority. The set of nodes $\Omega = \{\omega_0, \ldots, \omega_n\}$ can also be considered as a set of perspectives, labels or weights. Typically, there will be a unique top node, which will be called ω_0. Furthermore, all the information at ω_i has the same level of priority or certainty and has precedence over the information at ω_j where

$\omega_i > \omega_j$. When the knowledge base is totally ordered, we agree that the ordering is given by $\omega_0 > \omega_1 > \ldots > \omega_n$.

Some formalisms discussed here, interpret the default implication ($\rightarrow$) and the strict implication ($\Rightarrow$) as unidirectional, using modus ponens as inference rule. The consequence relation for these formalisms will be denoted by $\mid\!\sim$. Others however interpret the implications as material ones, using the classical consequence relation $\vdash$. As a result of this, they allow *contraposition*. In other words, from the default "birds fly" ($b \rightarrow f$), they conclude that nonfliers, by default, are not birds. It can be argued whether or not contraposition is a desirable property of a nonmonotonic formalism. E.g. [Gin94] for the default "humans tend not to be diabetics", it seems unreasonable to conclude from this that diabetics tend not to be human. If contraposition is not applicable in a formalism, it can be simulated by explicitly adding the contraposed information whenever required. However, when a formalism does allow contraposition, this reasoning mechanism cannot be disabled. The interpretation of implications and the allowance of contraposition are controversial topics on which the defeasible logics tend to disagree. Depending on the interpretation of rules, (in)consistency should be understood as (in)consistency with respect to the classical consequence relation $\vdash$, or to the restricted consequence relation $\mid\!\sim$.

An *interpretation* of a defeasible theory T is a function assigning a truth-value to each proposition occurring in T. An interpretation M is a *model* of T if it satisfies every observation and strict rule occurring in T, i.e. if $M \models o$ for each $o \in O$, and $M \models CB(r) \supset H(r)$ for each $r \in R_s$. Because a defeasible theory can contain defeasible and interfering rules which can be defeated, one can hardly expect that each rule is satisfied in a model. Indeed, a defeasible or interfering rule r can be verified, satisfied or falsified. A model M is said to *verify* r if $M \models CB(r) \wedge H(r)$, to *satisfy* r if $M \models CB(r) \supset H(r)$ and to *falsify* r if $M \models CB(r) \wedge \neg H(r)$.

In this chapter on defeasible logics the following formalisms will be discussed: basic defeasible logic [Nut92, Nut88], ordered logic [VNG89a, VNG90, GVN94, GLV91, Lae90], conditional entailment [GP92], Brewka's system of preferred subtheories [Bre89], Pearl's system Z [Pea90] and the argumentation-based system of Simari and Loui [SL92]. We will compare and catalogue these formalisms with respect to their basic design choices such as the approach to express priorities among defaults, the attitude towards conflicting defaults (skeptical or credulous), the interpretation of rules and the allowance of contraposition.

0.2 Pearl's system Z

Pearl's system Z [Pea90] is a defeasible logic formalism by means of which a total ordering can be imposed on most sets of defeasible rules. System Z is based on a probabilistic interpretation [Ada75, Pea89] of rules: a defeasible rule $A \rightarrow p$ is interpreted as asserting that the probability of p is high, given that A represents all the available evidence. As a result, system Z transforms a set of rules into a totally ordered partition $R_0, R_1, \ldots, R_n$, whenever possible. Such an ordered partition of a set of defeasible rules is called the Z-ordering. The idea is that lower ranked rules contain more normal, or less specific, information. The consequence relation of system Z is based on the preference relation among models [Sho87b, KLM90] which results from the rule ranking.

0.2.1 Deriving a natural ordering of defaults

In system Z, knowledge bases contain only defeasible rules. The default implication $\rightarrow$ is interpreted as material implication $\supset$. In other words, a rule

$$\{a_1, \ldots, a_n\} \rightarrow b$$

is treated as the logical formula

$$a_1 \wedge \ldots \wedge a_n \supset b$$

Therefore, no restrictions are imposed on models: because there are no strict rules which have to be satisfied, a model is an ordinary interpretation. The derivation of the Z-ordering is based on the notion of *toleration.*

Definition 1 Let R be a knowledge base containing only defaults. A subset $R' \subseteq R$ is said to *tolerate* a rule r if there is a model that verifies r and satisfies all rules in R'.

The process of finding a Z-ordering for a set of defaults R is defined as follows: every rule that is tolerated by all the other rules in R is in R_0. Next, every rule that is tolerated by the remaining ones (the rules in $R - R_0$) is in R_1. Continuing in this way, the process stops with a full partition, the Z-ordering, or with some rules which are not tolerated by the remaining ones. If there is a full partition, the set of defaults can be considered to be *Z-consistent.* Z-consistency is also called p-consistency in [Ada75] or ϵ-consistency in [Pea88]. For example, $R = \{a \rightarrow p, a \rightarrow \neg p\}$ is not Z-consistent, because none of the rules is tolerated by the other one.

System Z can only derive conclusions for Z-consistent sets of defaults. For such a set of defaults R, the ranking among rules can be translated into preferences among models. The rank of a rule r is given by $Z(r) = j$ iff $r \in R_j$. The rank associated with a model M is given by

$$Z(M) = min\{i \mid M \models CB(r) \supset H(r), Z(r) \geq i\}$$

which is the rank of the highest-ranked rule falsified by M plus 1. The rank of a formula f is given by

$$Z(f) = min\{Z(M) \mid M \models f\}$$

The next definition gives us a reasonable notion of entailment, based on the idea that a formula g is a plausible consequence of f if the models of $f \wedge g$ are preferred to the models of $f \wedge \neg g$.

Definition 2 Let R be a Z-consistent set of defaults and $O = \{o_1, o_2, \ldots, o_n\}$ a set of observations. A literal p is said to be *Z-entailed* [1] by $T = (O, R)$, denoted $T \vdash_z p$, if

$$Z(o_1 \wedge o_2 \wedge \ldots \wedge o_n \wedge p) < Z(o_1 \wedge o_2 \wedge \ldots \wedge o_n \wedge \neg p)$$

where $<$ is the Z-ordering induced by R.

As a result of this entailment definition, system Z can considered to be skeptical: whenever both formulae $o_1 \wedge o_2 \ldots \wedge o_n \wedge p$ and $o_1 \wedge o_2 \ldots \wedge o_n \wedge \neg p$ have the same Z-rank, no conclusion about p is possible.

In contrast to other formalisms based on probabilistic or preferential model ideas, like p-entailment [Ada75], ϵ-entailment [Pea88] and r-entailment [LM88], Z-entailment properly handles irrelevant features, e.g. from “birds fly” we can conclude that “red birds fly”. Z-entailment also sanctions rule chaining and allows contraposition, due to the classical logic interpretation of defaults. One of the shortcomings of system Z is that it suffers from the so-called *drowning problem*, as illustrated in the following example.

Example 1 Consider the knowledge base

$$R = \{p \rightarrow b, p \rightarrow \neg f, b \rightarrow f, b \rightarrow w, sf \rightarrow b\}$$

Let p stand for penguin, b for bird, f for fly, w for wings and sf for something feathered. The Z-ordering induced by R is

$$R_0 = \{b \rightarrow f, b \rightarrow w, sf \rightarrow b\},$$

$$R_1 = \{p \rightarrow \neg f, p \rightarrow b\}$$

When $O_1 = \{sf\}$, we get that $(O_1, R) \vdash_z f$, because $Z(sf \wedge f) = 0$ and $Z(sf \wedge \neg f) = 1$. This illustrates the fact that rules ($sf \rightarrow b$ and $b \rightarrow f$) can be chained. For $O_2 = \{\neg b\}$, system Z entails $\neg p$, because $Z(\neg b \wedge \neg p) = 0$ and $Z(\neg b \wedge p) = 2$. Therefore, we can conclude that contraposition (of the rule $p \rightarrow b$) is allowed. For the set of observations $O_3 = \{p, sf\}$, we get that $(O_3, R) \vdash_z \neg f$, because $Z(p \wedge sf \wedge \neg f) = 1$ and $Z(p \wedge sf \wedge f) = 2$. The ambiguity about f is solved correctly, based on specificity reasons: being a penguin is more specific than being a bird. The presence of the irrelevant feature sf causes no problem. However, w is not a Z-entailed conclusion:

$$Z(p \wedge sf \wedge w) = Z(p \wedge sf \wedge \neg w) = 1$$

Although a penguin is an exceptional bird with respect to the ability to fly, nothing prevents him from having wings. As a consequence of the Z-ordering, a penguin is declared to be an exceptional bird in all respects, so that no property of birds can be inherited. This weakness of system Z, which can also be found in some other

[1] Pearl uses the name 1-entailment instead of Z-entailment, and consistency instead of Z-consistency.

systems, is called the *drowning problem* [BCD+93]: the inability to sanction property inheritance from classes to exceptional sub-classes.

Another problem of this consequence relation is that the commitment to a unique integer ranking sometimes yields unintuitive results.

Example 2 Consider the knowledge base

$$R = \{gap \to p, p \to b, gap \to f, p \to \neg f, b \to f, s \to \neg f\}$$

together with the observations $O = \{gap, s\}$. Let gap stand for genetically-altered penguin, s for sick and let p, b and f be as in the previous example. The Z-ordering gives us $R_0 = \{b \to f, s \to \neg f\}$, $R_1 = \{p \to b, p \to \neg f\}$ and $R_2 = \{gap \to p, gap \to f\}$. Because $Z(gap \wedge s \wedge f) < Z(gap \wedge s \wedge \neg f)$, we get $(O, R) \vdash_z f$, which is not what we expect.

To remedy this kind of problems, a more refined ordering is required.

0.2.2 System Z^+: resolving remaining ambiguities by explicit means

System Z^+ [GP91] can be considered as an extension of system Z, evolved by the (correct) observation that not all priorities among rules are specificity-based. Therefore, there are priorities which cannot be extracted from the knowledge base, but should be encoded on a rule-by-rule basis. To make this possible, each default is supplied with an integer, signifying the strength with which the rule is stated. Similar to system Z, we want to make each model as normal as possible, by assigning to it the lowest possible non-negative integer permitted by the constraints. Once again, this ordering is unique. The process of finding this Z^+-ordering is slightly more complicated because of the presence of the strength associations, and computes the ranking of models and rules recursively in an interleaved fashion.

The step by step procedure for computing the Z^+ ranking for a set of defaults R and its models is defined as follows: Let R_0 be the set of rules tolerated by R. For each rule r_i with strength δ_i in R_0, set $Z^+(r_i) = \delta_i$. As long as there are rules without Z^+ rank, we can compute the Z^+ rank for models falsifying only rules having a Z^+ rank and verifying at least one of the other rules, by the formula

$$Z^+(M) = max\{Z^+(r_i) : M \models CB(r_i) \wedge \neg H(r_i)\} + 1$$

For each rule r_i without Z^+ rank which is verified in such a minimal model M, we can establish its Z^+ rank by

$$Z^+(r_i) = Z^+(M) + \delta_i$$

The definition for Z^+-entailment is similar to the one for Z-entailment.

The additional rule strengths make it possible to refine specificity-based priorities. The following remark can be made: no matter how we choose the integers assigned to the defaults, it is impossible to obtain $Z^+(r) < Z^+(r')$ whenever r contains more specific information than r'. In other words: the additional tools to encode explicit prioritization cannot override the specificity criterion, but can only help to resolve conflicts which are not specificity-based. The input priorities influence the ranking of rules, but don't dominate: they undergo adjustements so that compliance with specificity constraints is automatically preserved. However, some of the weaknesses of system Z are inherited, among which the inability to sanction inheritance across exceptional subclasses. The user can partially bypass this obstacle by means of the rule strengths he assigns to the involved rules. However, it is intuitively not clear [GP91] why strengths have to be assigned that way, and therefore, this solution is not entirely satisfactory.

0.3 Conditional entailment

In the system of conditional entailment [GP92, Gef92], some weaknesses of system Z are remedied. Instead of a deriving a total ordering on sets of defaults, an irreflexive and transitive (strict partial) order on defaults is extracted from the knowledge base. The notation $r < r'$ means that the default r' has higher priority than the default r. The admissible priority orderings should reflect the preferences implicit in the knowledge base. The resulting preference relation on models is partial as well, and favours models violating minimal sets of (low priority) defaults. Conditional entailment can be used for theories containing strict rules [2]

and defaults. Both kind of rules are interpreted as classical logic formulae, i.e. $\rightarrow$ and $\Rightarrow$ are treated as the material implication $\supset$.

Whether or not a priority ordering is called admissible is determined by the notion of conflict. A set of defaults $D \subseteq R_d$ is said to be *in conflict* with a default $r \in R_d$, in the context R, iff $B(r) \cup R_s \cup D \vdash \neg H(r)$, where $\vdash$ stands for the classical consequence relation, $B(r)$ and $H(r)$ for the body and the head of rule r, and R_s for the subset of R containing all the strict rules. The notion of conflict is related to the notion of toleration introduced by Pearl [Pea90]: a set of defaults D is in conflict with a rule r iff r is not tolerated by D.

Definition 3 Let R be a knowledge base containing strict and defeasible rules, in which R_d represents the subset of defeasible rules. An irreflexive and transitive ordering $<$ on R_d is an *admissible priority ordering* if every set of defaults $D \subseteq R_d$ in conflict with a default $r \in R_d$ contains a default $r' \in D$ such that $r' < r$.

The intuitive idea behind this definition is that when p is all the evidence that is given, a default $p \rightarrow q$ should be applied, even in the presence of sets of defaults in conflict with $p \rightarrow q$. A knowledge base R can have none, one or more admissible

[2] In Geffner's terminology, a set of strict rules L and a set of defaults D form together a background context. A default theory T consists of a background context and an evidence set E, containing information specific to the hand. A falsified default is called violated

priority orderings. Whenever there is at least one admissible priority ordering, R is called *conditionally consistent*. The concept of conditional consistency is similar to Z-consistency, p-consistency and ϵ-consistency.

Example 3 For a knowledge base

$$R = \{\neg p \rightarrow q, \neg q \rightarrow p, \rightarrow \neg p, \rightarrow \neg q\}$$

no admissible priority ordering can be found. This set is not conditionally consistent.

Usually, when M is a model of R, meaning that M satisfies all strict rules in R, some defeasible rules of R_d will be applicable, but not applied in M. These rules can be considered as falsified or violated rules, and will be denoted by $VS(M)$.

Definition 4 Let R be a knowledge base containing a set of strict rules R_s and a set of defeasible rules R_d. An *admissible prioritized structure* is a quadruple $(I, <_I, R_d, <)$, where I stands for the set of interpretations, $<$ is a priority ordering over R_d admissible with R and $<_I$ is a binary relation over I such that for two interpretations M and M', we have that $M <_I M'$ iff $VS(M) \neq VS(M')$ and

$$\forall r \in VS(M) - VS(M') \exists r' \in VS(M') - VS(M) \; : \; r < r'$$

Priority orderings may not contain infinite ascending chains $r_1 < r_2 < r_3 < \ldots$ With this restriction, it can be shown (see [Gef92]) that when $(I, <_I, R_d, <)$ is a prioritized structure, the pair $(I, <_I)$ is a preferential structure [KLM90, Mak89, Sho87a], meaning that the relation $<_I$ is also irreflexive and transitive. In a preferential structure $(I, <_I)$, $M <_I M'$ means that M is preferred to M'. M is a preferred model if there is no model M' preferred to M. Therefore, conditional entailment can be considered as an extension of preferential entailment, defined in terms of the class of admissible prioritized structures, induced by the admissible priority orderings on defaults. The following definition shows that the attitude towards conflicting defaults can be considered to be indirectly skeptical.

Definition 5 Let T be a defeasible theory consisting of the knowledge base R and a set of observations O. A literal p is *conditionally entailed* by T, denoted $T \vdash_{ce} p$, iff p holds in all the preferred models of T of every prioritized structure admissible with R.

It can be shown (see [Gef92]) that only minimal prioritized structures, induced by minimal admissible priority orderings, need to be considered. An admissible priority ordering is minimal when no set of tuples $r < r'$ can be deleted without violating the admissibility constraints. As a result, the cost of computing conditional entailment can be considerably reduced.

Another remark that can be made is that no admissible prioritized structures exist for conditionally inconsistent sets of rules. Because conditional entailment is defined in terms of such structures, it is restricted to be used for conditionally consistent sets of rules only.

Like system Z, conditional entailment sanctions rule chaining, contraposition and the discounting of irrelevant features. However, it doesn't suffer from the drowning problem and other problems caused by the commitment to a total ordering of defaults and models, as shown in the two following examples.

Example 4 Consider again the knowledge base [3]

$$R = \{p \to b, p \to \neg f, b \to f, b \to w, sf \to b\}$$

introduced in example 1. First we have to look for admissible priority orderings. The set of defaults $\{p \to b, b \to f\}$ is in conflict with $p \to \neg f$ because $\{p, p \supset b, b \supset f\} \vdash f$. Likewise, $\{b \to f, p \to \neg f\}$ is in conflict with $p \to b$. Because priority orderings are transitive and irreflexive, we obtain a unique (minimal) admissible priority ordering, where $b \to f < p \to \neg f$ and $b \to f < p \to b$. The priority of $p \to b$ over $b \to f$ is an important requirement caused by the allowance of contraposition. Indeed, without the priority $b \to f < p \to b$, we would not be able to conclude b from p. This can be considered as a slight disadvantage of the system, because this priority does not appear justified on specificity grounds. For a set of observations $O_1 = \{sf\}$, there is a class of models which violate no default, and which is therefore preferred. In these models, the literals sf, b, f, w and $\neg p$ hold, and are therefore conditionally entailed. These inferences involve default chaining ($sf \to b, b \to f$) and contraposition ($f \to \neg p$). It is obvious that $\neg p$ cannot be a conclusion in formalisms without contraposition. When an additional observation p is made, $O_2 = \{sf, p\}$ gives rise to three classes of minimal models. Models of the first class contain sf, p, b, w and $\neg f$ and violate default $b \to f$. Models of the second class contain sf, p, b, w and f and violate $p \to \neg f$. Models of the third class contain $sf, p, \neg b, \neg f$ and violate $p \to b$ and $sf \to b$. Because $b \to f < p \to \neg f$ and $b \to f < p \to b$, models of the first class are preferred, since they violate a less important default. Therefore, the literals b, w and $\neg f$ are conditionally entailed, illustrating that conditional entailment properly handles specificity information and doesn't suffer from the drowning problem.

This example also shows that conditional entailment can deal with irrelevant information: the fact that the penguin has feathers is irrelevant for concluding that he will not be able to fly.

Example 5 Reconsider the knowledge base

$$R = \{gap \to p, p \to b, b \to f, p \to \neg f, gap \to f, s \to \neg f\}$$

[3] In most formalisms allowing strict and defeasible rules, the knowledge that penguins are birds is expressed by a strict rule. To illustrate the impact of a set of defeasible rules, they usually give an alternative version of this sort of example, saying that "typical university-students are adults", "typical adults work" and "typical university-students don't work". However, as we pursue uniformity, we will adhere to the penguin example

introduced in example 2. The unique minimal admissible priority ordering is given by $p \to \neg f < gap \to p$, $p \to \neg f < gap \to f$, $b \to f < p \to b$, $b \to f < p \to \neg f$, $b \to f < gap \to p$ and $b \to f < gap \to f$.

As a result, there are two classes of minimal models for the set of observations $O = \{gap, s\}$: models of the first class contain the literals gap, s, p, b, f and models of the second class contain gap, s, p, b and $\neg f$. Both classes are preferred. Therefore, no conclusion can be made about f or $\neg f$, which corresponds to our intuition.

Unfortunately, it can be shown that conditional entailment has a problem with inheritance reasoning, as a result of allowing contraposition:

Example 6 Consider the knowledge base [GP92]

$$R = \{a \to b, b \to c, c \to \neg d, a \to d\}$$

There are four admissible and minimal priority orderings. When the observation a is made, all four priority orderings lead to the same two classes of minimal models. Models of the first class contain a, b, c and d, while models of the second class contain a, b, d and $\neg c$. This last class of models doesn't occur in formalisms without contraposition. We get that a, b and d are conditionally entailed. However, c is not conditionally entailed, whereas it would be sanctioned by most inheritance reasoners.

This example illustrates that conditional entailment does not subsume inheritance reasoning. This problem is acknowledged by Geffner [GP92], but no solution is proposed. System Z suffers from the same problem.

In [GP92, Gef92], Geffner also proposes a proof theory for computing the conditionally entailed literals. The proof theory is structured around the notion of arguments [Lou87, Pol87a] and uses the admissible priority orderings on rules to select arguments.

0.4 The argument-based system of Simari and Loui

Simari and Loui present an argumentative approach [SL92] to defeasible reasoning. An argument can be considered as a set of defaults indicating support for a certain literal. However, this support doesn't guarantee that the literal will be concluded: counterarguments and specificity should also be taken into account. The system of Simari and Loui combines features of prominent argument-based formalisms: the notion of argument of Loui [Lou87], the specificity comparator of Poole [Poo85] and the interaction among arguments as described by Pollock in his theory of warrant [Pol87b]. The problem with Poole's specificity comparator is that nothing is said about how to apply it to interactions among arguments. Pollock on the other hand treats the interaction among arguments properly, but doesn't rely on specificity. The early definition of Loui was insufficient from a mathematical point of view.

The defeasible theories [4] which are considered in this system usually contain a set of defaults R_d, a set of strict rules R_s representing necessary information and a set of observations O. The rules can contain free variables, but arguments are composed using grounded instances. Consequences are derived using modus ponens and, in the first order case, instantiation. Contraposition is not allowed. The only condition on a defeasible theory containing rules R and observations O is that $O \cup R_s$ is consistent. In what follows, we consider strict and default rules to be grounded. Furthermore, observations are grounded literals.

In Simari's and Loui's argument-based system, an argument structure is a consistent set of defaults needed to derive a literal. However, in contrast to Poole's definition [Poo85], this set of defaults needs to be minimal. Such a minimal set does not contain a redundant rule, i.e. a rule that is unnecessary for infering the literal. Similar to Poole, an irreflexive an transitive priority relation on argument structures is derived, based on specificity.

Definition 6 Let O be a set of observations, and R a knowledge base with strict part R_s and defeasible part R_d in the defeasible theory (O, R). Let $A \subseteq R_d$ be a set of defaults and p a literal. (A, p) is an *argument structure* iff A is a minimal set of defaults for which $O \cup R_s \cup A \mid\!\sim p$ and $O \cup R_s \cup A$ is consistent. For two argument structures (A, p) and (B, q) such that $B \subseteq A$, we say that (B, q) is a *subargument* of (A, p), denoted $(B, q) \subseteq (A, p)$.

Definition 7 Let $T = (O, R)$ be a defeasible theory and let (A_1, p_1) and (A_2, p_2) be two argument structures. The argument structure (A_1, p_1) is said to be *strictly more specific* than the argument structure (A_2, p_2) iff

1. for each set of grounded literals E such that $R_s \cup E \cup A_1 \mid\!\sim p_1$ and $R_s \cup E \not\mid\!\sim p_1$, it is also the case that $R_s \cup E \cup A_2 \mid\!\sim p_2$; and

2. $\exists E$ such that $R_s \cup E \cup A_2 \mid\!\sim p_2$, $R_s \cup E \not\mid\!\sim p_2$ and $R_s \cup E \cup A_1 \not\mid\!\sim p_1$.

The idea behind this definition is that argument A_1 is strictly more specific than argument A_2 if every nontrivial condition which activates A_1, also activates A_2, but not the other way round. This priority ordering on argument structures can then be used to solve conflicts, by selecting those argument structures which are "better" than others. For this purpose, some relations among argument structures need to be defined.

Definition 8 Let $T = (O, R)$ be a defeasible theory and let (A_1, p_1) and (A_2, p_2) be two argument structures. The argument structures (A_1, p_1) and (A_2, p_2) are said to *disagree* iff $R_s \cup O \cup \{p_1, p_2\}$ yields an inconsistency. The argument structure (A_1, p_1) is a *counterargument* of (A_2, p_2) at p iff there is a subargument (A, p) of (A_2, p_2) such

[4] In Simari's and Loui's terminology, a defeasible theory is called a defeasible logic structure. Such a structure contains necessary and contingent facts, corresponding to strict rules and observations, together with a set of defeasible rules.

that (A_1, p_1) and (A, p) are in disagreement with each other. The argument structure (A_1, p_1) *defeats* (A_2, p_2) iff there is a subargument (A, p) of (A_2, p_2) such that (A_1, p_1) counterargues (A_2, p_2) at p, and (A_1, p_1) is strictly more specific than (A, p).

For a certain argument structure (A, p), there can be a set of argument structures interfering with (A, p), i.e. counterargueing (A, p). Among those interfering argument structures, there may be some defeaters of (A, p), which could in turn be defeated. When no defeater remains undefeated, (A, p) is reinstated. This inductive approach is based on Pollock's method [Pol87b] of defining which arguments survive counterarguments, although he uses only one kind of label, whereas Simari and Loui use two: an S-label to indicate support, and an I-label to indicate interference.

Definition 9 An argument structure can be an S^j or I^j argument, where S stands for supporting argument, I for interfering argument, and j is the level at which the argument is supporting or interfering.

1. each argument structure is an S^0 and an I^0 argument.

2. (D_1, p_1) is an S^{n+1} argument iff $\nexists I^n$ argument (D_2, p_2) such that (D_2, p_2) is a counterargument of (D_1, p_1) at some p.

3. (D_1, p_1) is a I^{n+1} argument iff $\nexists I^n$ argument (D_2, p_2) such that (D_2, p_2) defeats (D_1, p_1).

The idea behind this definition is that an argument structure retains its interfering capacity as long as it is not defeated.

Definition 10 An argument structure (D, p) *justifies* p iff $\exists m$ such that $\forall n \geq m$, (D, p) is an S^n argument for p. A literal p is an *argument-based consequence* of a defeasible theory T, denoted $T \vdash_{abs} p$, if there is an argument structure (D, p) justifying p.

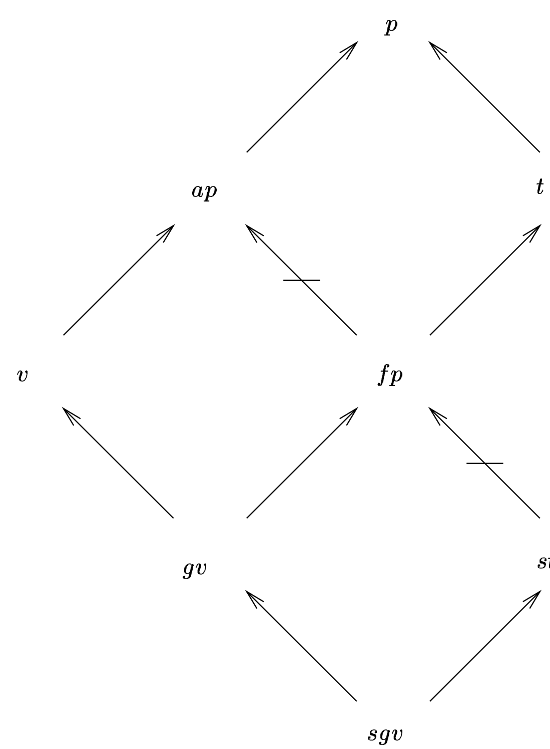
This theory of justifying argument structures is well-behaved: it can be shown [SL92] that a subargument of a justifying argument structure is a justifying argument structure as well. In other words, when (A_1, p_1) and (A_2, p_2) are two argument structures such that $(A_2, p_2) \subseteq (A_1, p_1)$ and (A_1, p_1) justifies p_1, then also (A_2, p_2) justifies p_2.

The argument-based system properly handles the examples 1, 2 and 6. More specifically, it doesn't suffer from the problems encountered by adhering to a total ordering or allowing contraposition. Furthermore, the system is not restricted to consistent sets of defaults, as is the case for system Z (Z-consistent sets) and conditional entailment (conditionally consistent sets).

By making the distinction between supporting and interfering arguments and considering an inductive definition of justification, the skeptical attitude towards conflicting defaults turns out to be ambiguity-propagating. However, it can be shown that an ambiguity-propagating skeptical formalism such as the argument-based system of

Simari and Loui is still not ideally skeptical: sometimes, a literal which holds in every possible world, is not directly derivable. We will illustrate both aspects of this skeptical approach by means of an example given by Stein [Ste92].

Example 7 Consider the set of rules $R = \{sgv \rightarrow gv, sgv \rightarrow st, gv \rightarrow fp, st \rightarrow \neg fp, gv \rightarrow v, v \rightarrow ap, fp \rightarrow \neg ap, fp \rightarrow t, ap \rightarrow p, t \rightarrow p\}$, together with the observation set $O = \{sgv\}$. Let sgv stand for seedless grape vine, gv for grape vine, st for seedless thing, fp for fruit plant, v for vine, ap for arbor plant, t for tree and p for plant. This defeasible theory is based on information which can be represented in the following inheritance network.



For this defeasible theory, we get the following argument structures:

$$
\begin{aligned}
A_0 &= (\emptyset, sgv) \\
A_1 &= (\{sgv \rightarrow gv\}, gv) \\
A_2 &= (\{sgv \rightarrow st\}, st) \\
A_3 &= (\{sgv \rightarrow gv, gv \rightarrow fp\}, fp) \\
A_4 &= (\{sgv \rightarrow st, st \rightarrow \neg fp\}, \neg fp) \\
A_5 &= (\{sgv \rightarrow gv, gv \rightarrow v\}, v) \\
A_6 &= (\{sgv \rightarrow gv, gv \rightarrow v, v \rightarrow ap\}, ap) \\
A_7 &= (\{sgv \rightarrow gv, gv \rightarrow fp, fp \rightarrow \neg ap\}, \neg ap) \\
A_8 &= (\{sgv \rightarrow gv, gv \rightarrow fp, fp \rightarrow t\}, t)
\end{aligned}
$$

$$A_9 = (\{sgv \to gv, gv \to fp, fp \to t, t \to p\}, p)$$
$$A_{10} = (\{sgv \to gv, gv \to v, v \to ap, ap \to p\}, p)$$

No argument structures counterargue A_0, A_1, A_2 and A_5, and the literals sgv, gv, st and v are argument-based consequences. Although the argument structure A_4 is a counterargument for A_7 at literal fp, A_7 is not defeated by A_4. The reason for this is that none of the two counterargueing argument structures A_3 and A_4 is more specific than the other one. As a result, ap is not an argument-based consequence: A_6 is an argument for ap, but A_7 is an argument for $\neg ap$ which preserves its interfering capacity, even though it is based on the ambiguous proposition fp. The ambiguity about fp is propagated forwards, allowing it to interfere with the possible derivation of ap. In an ambiguity-blocking skeptical formalism on the other hand, the two rules $gv \to fp$ and $st \to \neg fp$ would defeat each other, the ambiguity about fp would be forgotten and the rule $v \to ap$ would be left unchallenged. Unfortunately, we can also show that p is not an argument-based consequence: A_9 is counterargued by the undefeated argument structure A_4 at proposition fp, while A_{10} is counterargued by the undefeated argument structure A_7 at proposition ap. However, p is in every possible world or credulous extension, so that the system is not ideally skeptical. Summarizing this discussion, we can consider the levels of support and interference, resulting in the following table:

Level	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
0	IS	IS	IS	IS	IS	IS	IS	IS	IS	IS	IS
1	IS	IS	IS	I	I	IS	I	I	I	I	I
2	IS	IS	IS	I	I	IS	I	I	I	I	I

As a result, the argument-based consequences of this default theory are the literals justified by the argument structures A_0, A_1, A_2 and A_5: sgv, gv, st and v.

Furthermore, Simari and Loui claim that only strict specificity is used to derive the ordering on argument structures, but this is not entirely true: besides strict rules, the defeasible rules contained in the argument structure, and therefore a minimum of defeasible information, is used.

Example 8 Consider the knowledge base

$$R = \{p \to b, b \to f, p \to \neg f\}$$

together with the observations $O_1 = \{p, b\}$. There is no specificity between the argument structures $(\{p \to \neg f\}, \neg f)$ and $(\{b \to f\}, f)$, so that the argument-based consequences are p and b. When we only have p as observation, i.e. $O_2 = \{p\}$, we have that the argument structure $(\{p \to \neg f\}, \neg f)$ is more specific than $(\{p \to b, b \to f\}, f)$, so that the first argument structure defeats the second one and the argument-based consequences are p, b and $\neg f$. Indeed, in the first case the default rule $p \to b$ is not used to derive specificity, as it is not contained in the second argument structure because of the minimality requirement. However, in the second case it is part of the argument for f, and therefore, specificity solves the conflict.

This discrepancy, caused by explictly adding an observation which can be explained directly by the knowledge base, might not be what we expect.

A more severe problem arises from the fact that only a minimum of defeasible information is used to determine specificity: in some cases, the system cannot properly handle irrelevant features, as shown in the following example.

Example 9 When we augment the set of rules of the previous example with the rule $sf \rightarrow b$ ("Typically, something feathered is a bird") and we consider the observation set $O = \{sf, p\}$, we get 3 argument structures about f: $A_1 = (\{p \rightarrow \neg f\}, \neg f)$, $A_2 = (\{p \rightarrow b, b \rightarrow f\}, f)$ and $A_3 = (\{sf \rightarrow b, b \rightarrow f\}, f)$. The argument structure A_1 defeats A_2, but A_3 keeps its interfering capacity, so that A_1 doesn't become reinstated. Therefore, nothing can be concluded about f, while intuitively we would expect f, as sf is irrelevant with respect to f.

Several other formalisms for defeasible reasoning [Poo85, Vre91, Pol92, GV95] are based on arguments. However, the structure of arguments can differ. E.g. in the formalism of Vreeswijk [Vre91], strict rules and observations can be part of an argument. In this formalism, deductive arguments (arguments based on standard propositional logic) defeat arguments involving defaults. An argument based on a single default defeats an argument based on two or more defaults. Once again, the intuitive idea here is that when p is all the evidence that is given, a default $p \rightarrow q$ should be applied, even in the presence of sets of defaults in conflict with $p \rightarrow q$. The only other possibility for an argument A_1 to defeat a conflicting argument A_2 is that both A_1 and A_2 are based on a single default, where A_1 is based on the *most specific reference class*. In other words, the antecedent of the default used in A_2 should deductively follow from the antecedent of the default used in A_1. For other conflicting arguments, no kind of specificity is considered in an attempt to resolve ambiguities. Since the formalism is credulous, this approach usually yields a wide number of extensions or possible worlds. In [Dun95], a more general method of accepting arguments, which are treated as abstract entities, is described.

0.5 Brewka's preferred subtheories

The system of preferred subtheories of Brewka [Bre89] deals with ordered knowledge bases and is based on the notion of preferred maximal consistent subsets, as defined by Rescher [Res64]. Knowledge bases in this formalism can be considered as sets of defaults, interpreted as material implications. No additional set of observations is needed here, as observations can be integrated in a knowledge base by means of defeasible rules with empty bodies. The first version of this formalism can be considered as a generalization of Poole's Theorist approach [Poo88], where the prioritization is given by a total ordering on sets of rules. Instead of using two levels of formulae, called facts and defaults, more levels are allowed. Furthermore, even the most reliable formulae may be defeated. The idea is to use the total ordering on sets of defaults to prefer some maximal consistent subsets, namely the ones containing the most reliable information. When there is conflicting information with the same reliability, the attitude is credulous, so that several subsets can be selected. These selected subsets will be called preferred subtheories (or subbases).

Definition 11 Let $T = (\Omega, \leq, R, f)$ be a totally ordered theory. A maximal consistent subset $R' \subseteq R$ is a *preferred subtheory* iff $\forall k, 0 \leq k \leq n, R'$ contains a maximal consistent subset of $f(\omega_0) \cup \ldots \cup f(\omega_k)$.

In other words, a preferred subtheory of T can be obtained by starting with any maximal consistent subset of $f(\omega_0)$, adding as many formulas from $f(\omega_1)$ as possible, with respect to consistency, and continuing this process up till $f(\omega_n)$. Note that default implications are interpreted as material ones, so that consistency must be interpreted with respect to the classical consequence relation $\vdash$.

Once the preferred subtheories are known, the consequence relation can be defined in a credulous way, saying that p is a consequence if there is a preferred subtheory S such that $S \vdash p$, or in a skeptical way, saying that p is a consequence if $S \vdash p$ for all preferred subtheories S. The credulous approach yields several extensions, corresponding to the preferred subtheories. Because the skeptical consequence relation is not defined in a direct way but by detouring through the set of all preferred subtheories, this approach is indirectly skeptical. [5]

Definition 12 Let T be a totally ordered theory. A literal p is a *preferred consequence*, denoted $T \vdash_{ps} p$, iff $S \vdash p$ for each preferred subtheory S of T. The preferred extension is given by

$$E_{ps}(T) = \{p \mid T \vdash_{ps} p\}$$

Although the indirectly skeptical approach to the system of preferred subtheories gives results corresponding to our intuition, it also has a drawback: the number of preferred subtheories which have to be considered can become very large, so that checking whether or not a literal is a preferred consequence can become computationally costly. A suggestion to reduce the number of preferred subtheories which have to be investigated is given by the Lex consequence relation [BCD+93] where cardinality is used as a selection tool.

Definition 13 Let $T = (\Omega, \leq, R, f)$ be a totally ordered theory. $R' \subseteq R$ is a *Lex-preferred subtheory* iff R' is a preferred subtheory and for each preferred subtheory R^*, $\nexists i$ such that

$$|R' \cap f(\omega_i)| < |R^* \cap f(\omega_i)|$$

[5] The credulous and (indirectly) skeptical consequence relations are referred to as weak and strong provability by Brewka.

and

$$\forall j < i \;:\; |R' \cap f(\omega_j)| = |R^* \cap f(\omega_j)|$$

A literal p is a *Lex consequence*, denoted $T \vdash_{Lex} p$, iff $S \vdash p$ for each Lex-preferred subtheory S of T. The Lex extension is given by $E_{Lex}(T) = \{p | T \vdash_{Lex} p\}$.

It is obvious [BDP93] that whenever a literal is a preferred consequence, it is also a Lex consequence, because each Lex-preferred subtheory is a preferred subtheory. Although this proposal seems to be an interesting attempt to reduce the cost of computing whether or not a literal should be entailed, the Lex consequence relation sometimes gives unintuitive results, as is illustrated in the following example.

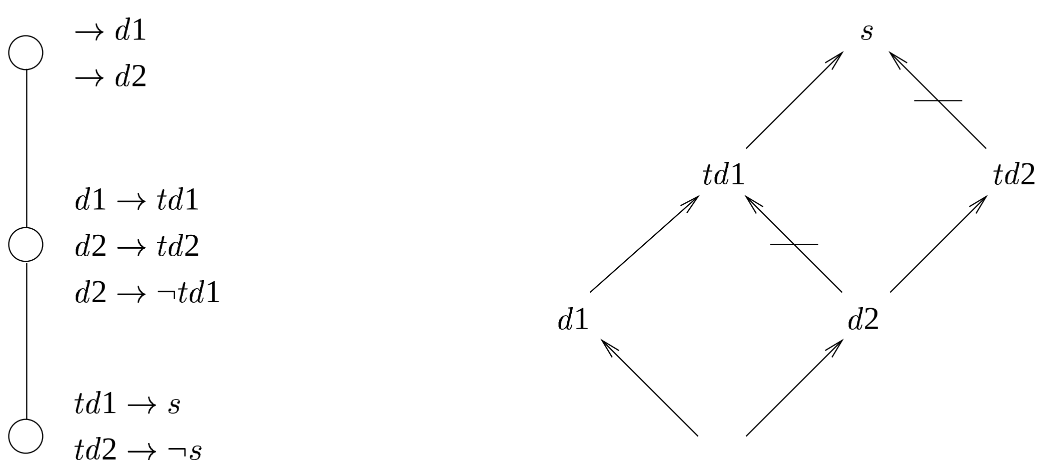
Example 10 Consider the totally ordered theory

$$\begin{aligned} T &= (\{\omega_0, \omega_1, \omega_2\}, \leq, \{\rightarrow d1, \\ &\rightarrow d2, d1 \rightarrow td1, d2 \rightarrow td2, d2 \rightarrow \neg td1, td1 \rightarrow s, td2 \rightarrow \neg s\}, f) \end{aligned}$$

where $\omega_0 > \omega_1 > \omega_2$,

$$\begin{aligned} f(\omega_0) &= \{\rightarrow d1, \rightarrow d2\} \\ f(\omega_1) &= \{d1 \rightarrow td1, d2 \rightarrow td2, d2 \rightarrow \neg td1\} \\ f(\omega_2) &= \{td1 \rightarrow s, td2 \rightarrow \neg s\} \end{aligned}$$

This theory can be considered as a variation on the extended Nixon-diamond [HTT87, THT87], as given in the context of semantic networks. Both the ordered theory and the semantic network are given in the figure below.



T Γ

The interpretation given to the example is the following: For $disease_1$ ($d1$), it is a good thing to take $drug_1$ ($td1$). On the other hand, when a person has $disease_2$ ($d2$), $drug_2$ ($td2$) could probably help, but taking $drug_1$ could make things worse. For $drug_2$, no side effects can be shown, and taking this drug normally doesn't make a patient sleepy (s). $Drug_1$ however can make a patient sleepy. John is very unlucky, because he has both $disease_1$ and $disease_2$, and doesn't know what to do.

There are three preferred subtheories, given by

$$\begin{aligned} PS_1 &= \{\rightarrow d1, \rightarrow d2, d1 \rightarrow td1, d2 \rightarrow td2, td1 \rightarrow s\} \\ PS_2 &= \{\rightarrow d1, \rightarrow d2, d1 \rightarrow td1, d2 \rightarrow td2, td2 \rightarrow \neg s\} \\ PS_3 &= \{\rightarrow d1, \rightarrow d2, d2 \rightarrow td2, d2 \rightarrow \neg td1, td1 \rightarrow s, td2 \rightarrow \neg s\} \end{aligned}$$

Only PS_3 is also Lex-preferred. As a result, following an indirectly skeptical approach, the preferred extension is given by $E_{ps}(T) = \{d1, d2, td2\}$, and the Lex extension by $E_{lex}(T) = \{d1, d2, \neg td1, td2, \neg s\}$. The preferred extension corresponds to our intuition and illustrates the fact that the system of preferred subtheories has no problem with propagating ambiguities: the ambiguity about $td1$ is propagated forwards, allowing it to interfere with the derivation of $\neg s$. The extension obtained by the Lex consequence relation seems unacceptable for this example: taking $drug_1$ is a possibility which should not be totally excluded. Whether John has to take $drug_1$ depends on the seriousness of the case and the judgement of the doctor.

Realizing that it might be difficult or even impossible sometimes to decide whether a rule r is of more, less or the same reliability as another rule r', Brewka proposes a second generalization in which it is possible to deal with partially ordered knowledge bases. In this second version, a strict partial order is given on the set of rules R. Again, preferred subtheories can be defined based on this ordering.

Definition 14 Let R be a finite set of rules and $<$ a strict partial ordering on R. A maximal consistent subset $R' \subseteq R$ is a *preferred subtheory* if there is a strict total ordering $(r_1, r_2, \ldots, r_l)$ of R respecting $<$ such that $R' = R'_l$ with $R'_0 = \emptyset$ and $R'_{i+1} = R'_i \cup \{r_{i+1}\}$ when r_{i+1} is consistent with R'_i, or $R'_{i+1} = R'_i$ otherwise.

It is obvious that we can translate such a set of partially ordered rules $(R, <)$ into a general ordered theory by creating a unique node for each rule, containing this rule, and by translating the partial order on rules into one on nodes.

0.6 Ordered logic and basic defeasible logic

Basic defeasible logic [Nut92, Lae90, GVN94, Gee96] and ordered logic [Lae90, GVN94, Gee96] are two related formalisms: both are directly skeptical approaches with a proof theory based on the concept of a proof tree. A proof tree contains positive conclusions for derivable formulae and negative conclusions for demonstrably non-derivable formulae. Negative conclusions are needed to show that a rule can only be applied whenever no potential defeater is applicable. Nute's early work on defeasible logic [Nut85, Nut88] was designed to be used for a single-perspective or single agent system, and is based on an implicit partial ordering on rules, resulting from specificity information. As a response to this early work, ordered logic presented a formalism suited to be used in multi-perspective or multi-agent environments [VNG89b], by structuring rules in an explictly given partially ordered set of perspectives. In general, a formalism with an explicit means to express priorities can be seen as a generalization of a related formalism using implicit specificity, because not all priorities are specificity-based. An additional advantage of ordered logic is that the ordering is on perspectives, i.e. on sets of rules, instead of on rules, making ordered logic well suited for reasoning with multiple perspectives or multiple agents. The family of defeasible logics presented in [Nut92] contains a variant on Nute's original defeasible logic where an explicit ordering is given, but here the ordering is on rules instead of on sets of rules. Regardless of the origin of the partial order, the idea that is used in the proof theory remains the same: an applicable rule will be applied only if every competing rule is either weaker, or can be shown to be not applicable. Both formalisms interpret rules in a unidirectional way, avoiding contraposition. Their skeptical character turns out to be ambiguity-blocking: as soon as an ambiguity cannot be solved by the priority ordering, they forget about it, so that it cannot interfere with other conclusions. For ordered logic, all rules are defeasible. Rules corresponding to observations and strict rules are simply assigned to a high priority perspective. Besides default rules, Nute's defeasible logic also takes observations, strict rules and defeaters into account. Defeaters never directly support conclusions, but can defeat rules that otherwise might be applied. The rule "Something that looks red under red light might not be red" is an example of a defeater. Except for some minor differences, mainly caused by the presence of strict rules and defeaters, the proof theory of the explicit version of Nute's defeasible logic is similar to the proof theory of ordered logic.

0.6.1 Implicit version of Nute's basic defeasible logic

We will start with describing Nute's earliest work on defeasible logic [Nut86, Nut88, Nut90] in which specificity is used to derive an implicit partial order on rules. Although in this early work, proofs are sequentially structured, we will present here a modified version [Nut92] where the proof theory is based on the concept of a proof tree, clarifying the structure of the proof. A default theory following Nute can contain strict rules, defeasible rules, [6] defeaters and observations.

The *evidentiality* symbol E is used to create E-sentences: where p is a literal, Ep can be read as "Evidently, p". A sentence is a literal or an E-sentence. The inference mechanism will avoid using rules that may later turn out to be defeated by restricting application to those rules that will definitely be not defeated. For any rule that is applied,

[6] Nute uses the symbols $\rightarrow$ and $\Rightarrow$ in the opposite way: in his logics, $\rightarrow$ indicates a strict rule and $\Rightarrow$ a defeasible rule.

the proof theory will first require that we show that no potential defeater is applicable. This is done by not only inferring positive conclusions like “p holds”, denoted as p^+, but also negative ones like “demonstrably, p does not hold”, denoted as p^-. A rule r_1 can only be defeated by a conflicting rule r_2 when r_1 is not superior to r_2. The idea is that every strict rule is superior to every defeasible or interfering rule. For the set of defeasible and interfering rules, one rule is considered to be superior to another if the antecedent of the first rule is more specific than the antecedent of the second one. Originally, Nute allows only strict rules to determine the specificity relations among antecedents. This kind of specificity is called *strict specificity*. Following the idea of strict specificity, a defeasible rule $A \to p$ is said to be superior to another defeasible rule $B \to \neg p$, iff for each $b \in B$, a proof exists for b^+, given the observation set A, and for some $a \in A$, a proof exists for a^-, given B. The resulting defeasible logic will be called $BDL_{\Rightarrow}$, indicating that only strict rules are used to uncover the implicit specificity information. With the defeasible logic $BDL_{\Rightarrow}$, several examples of nonmonotonic reasoning can be solved correctly. However, strict specificity is not sufficient: sometimes defeasible rules should also be used in determining specificity, as illustrated in the following example.

Example 11 Consider the defeasible theory T with rules

$$R = \{p \to b, p \to \neg f, b \to f\}$$

and observation set $O = \{p\}$. Using $BDL_{\Rightarrow}$, based on strict rules, we cannot show that $p \to \neg f$ is superior to $b \to f$, while intuitively, this is what we want.

This deficiency is dealt with by Nute [Nut92] by introducing what is called *defeasible specificity*, meaning that strict and defeasible rules are used to determine specificity. The basic idea is to simply adopt the proof theory of $BDL_{\Rightarrow}$ such that a defeasible rule $A \to p$ is superior to a rule $B \to \neg p$ just in case that for each $b \in B$ there is a proof for Eb^+, given the observation set A and there is a proof for Ea^-, starting from B, for some $a \in A$. But the following example given by Nute [Nut92] illustrates that this approach is not quite right.

Example 12 Consider the defeasible theory T with rules

$$R = \{b \Rightarrow m, b \to f, m \to \neg f, \{m, s\} \to b, \to s\}$$

and observation set $O = \{b\}$. This default theory can be interpreted as follows: bats (b) are mammals (m), bats normally fly (f), mammals normally do not fly, mammals with a sonar (s) are normally bats, and we consider a particular bat which, presumably, has a sonar. It is possible to construct a proof for Em^+ based on $\{b\}$ using the criterion of defeasible specificity just described. However, we can also construct a proof for Eb^+ based on $\{m\}$ by using the defeasible rule $\to s$. For this theory, we get that the bat is a mammal, but also that the mammal is a bat, so that neither $b \to f$ nor $m \to \neg f$ can be shown to be superior to the other one.

In order to solve this kind of problems, Nute limits the rules that can be used in determining defeasible specificity to the strict rules, the interfering rules, and the defeasible rules with non-empty antecedents. Let $R_{d_\emptyset}$ denote this subset of defeasible rules with non-empty antecedents, i.e.

$$R_{d_\emptyset} = R_d - \{\to p\ : \to p \in R_d\}$$

The resulting defeasible logic will be called $BDL_{\to}$, referring to the use of defeasible specificity [7]. In the proof theory, nodes get labels composed of three components: the first one represents what is being proved, the second one the set of literals which we consider to be true (the observations), and the third one the set of defeasible rules under consideration. The observational component is needed to show that one rule is superior to another one. The set of defeasible rules integrated into a label will mostly be the original set of defeasible rules, but it can be reduced to the subset of defeasible rules with non-empty antecedent, when specificity is to be determined.

Definition 15 Let $T = (O, R)$ be a defeasible theory. Where p is a sentence (i.e. a literal or an E-sentence) and s is $+$ or $-$, a *proof tree* for p^s in T using $BDL_{\to}$ is a finite tree where each node is labeled (q^t, E, D), where q is a sentence, t is $+$ or $-$, E is a set of literals and D is a set of defaults such that the root is labeled (p^s, O, R_d) and each node m satisfies one of the following conditions:

(D1) m is labeled (q^+, E, D) and either $q \in E$ or there is a strict rule $A \Rightarrow q \in R_s$ such that for each $a \in A, m$ has a child node labeled (a^+, E, D).

(D2) m is labeled (q^-, E, D), $q \notin E$, and for every strict rule $A \Rightarrow q \in R_s$, there is $a \in A$ and a child node of m labeled (a^-, E, D).

(D3) m is labeled (Eq^+, E, D) and m has a child node labeled (q^+, E, D).

(D4) m is labeled (Eq^+, E, D) and m has a child node labeled ($\neg q^-, E, D$) and there is a strict rule $A \Rightarrow q \in R_s$ such that for each $a \in A, m$ has a child node labeled (Ea^+, E, D) and for each $B \Rightarrow \neg q \in R_s$, there is $b \in B$ and a child node of m labeled (Eb^-, E, D).

(D5) m is labeled (Eq^+, E, D) and m has a child node labeled ($\neg q^-, E, D$) and there is a defeasible rule $A \to q \in D$ such that

[7] In [Nut92], where defeasible logics are represented as sets of conditions which have to be satisfied by each node in a proof tree, $BDL_{\Rightarrow}$ is originally called $SD_{\to}$, standing for the set of conditions

$$\{M^+, M^-, E^+, E^-, SS^+, D^+_{\to}, SD^-_{\to}\}$$

$BDL_{\to}$ is originally called $SD_{\Rightarrow}$, standing for the set of conditions

$$\{M^+, M^-, E^+, E^-, SS^+, D \Rightarrow^+, SD \Rightarrow^-\}$$

1) for each $a \in A, m$ has a child node labeled (Ea^+, E, D),
2) for each strict rule $B \Rightarrow \neg q \in R_s$, there is $b \in B$ and a child node of m labeled (Eb^-, E, D), and
3) for each defeasible rule $C \rightarrow \neg q \in D$ or interfering rule $C \rightsquigarrow \neg q \in R_i$, either
   a) there is $c \in C$ and a child node of m labeled (Ec^-, E, D), or
   b) for each $c \in C$, there is a child node of m labeled $(Ec^+, A, R_{d\emptyset})$ and for some $a \in A$, there is a child node of m labeled $(Ea^-, C, R_{d\emptyset})$.

(D6) m is labeled (Eq^-, E, D), m has a child node labeled (q^-, E, D) and m has a child node labeled $(\neg q^+, E, D)$.

(D7) m is labeled (Eq^-, E, D)

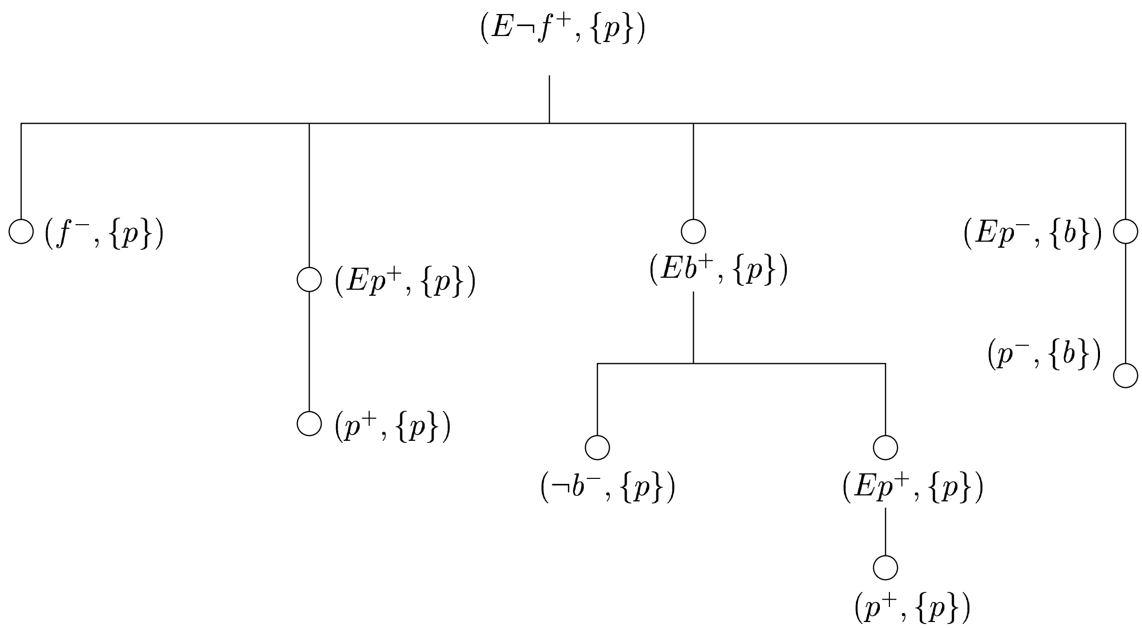
1) m has a child node labeled (q^-, E, D),
2) for each strict rule $A \Rightarrow q \in R_s$, either
   a) there is $a \in A$ and a child node of m labeled (Ea^-, E, D), or
   b) there is $B \Rightarrow \neg q \in R_s$ such that for each $b \in B, m$ has a child node labeled (Eb^+, E, D), and
3) for each defeasible rule $A \rightarrow q \in D$, either
   a) there is $a \in A$ and a child node of m labeled (Ea^-, E, D), or
   b) there is a strict rule $B \Rightarrow \neg q \in R_s$ such that for each $b \in B, m$ has a child node labeled (Eb^+, E, D), or
   c) there is a defeasible rule $C \rightarrow \neg q \in D$ or an interfering rule $C \rightsquigarrow \neg q \in R_i$ such that for each $c \in C$, m has a child node labeled (Ec^+, E, D) and either for each $a \in A, m$ has a child node labeled $(Ea^+, C, R_{d\emptyset})$, or there is $c \in C$ and a child node of m labeled $(Ec^-, A, R_{d\emptyset})$.

Condition (D1) captures the monotonic derivability of a literal, while (D2) shows when a literal is demonstrably not monotonically derivable. (D3) says that any literal that is monotonically derivable is also evidently the case. (D6) expresses that a literal can be shown to be not evident when its complement is monotonically derivable, unless the literal itself is also monotonically derivable. Conditions (D1), (D2), (D3) and (D6) can be considered as the monotonic kernel of the defeasible logic $BDL_{\rightarrow}$. Condition (D4) says that the consequent of a strict rule is evident if its antecedent is evident, the complement of its consequent is not strictly derivable and the rule is not defeated by another strict rule. Condition (D4) could be replaced by a stronger one [Nut92], where strict rules are used "more strictly" than here, in the sense that they are not allowed to defeat each other. (D5) says that the consequent of a defeasible rule is evident if the complement of its consequent is not strictly derivable, its antecedent is evident, the rule is not defeated by a strict rule, and the rule is superior to any other defeasible or interfering conflicting rule which cannot be shown to be non-applicable. (D7) says that a literal is demonstrably not evident if the literal is demonstrably not monotonically derivable and every rule which could derive the literal can be shown to be non-applicable or defeated.

Definition 16 Where T is a defeasible theory and p a sentence (i.e. a literal or an E-sentence), p is *$BDL_{\rightarrow}$-derivable* from T, denoted $T \vdash_{BDL_{\rightarrow}} p^+$, if there is a proof tree for p^+ in T using $BDL_{\rightarrow}$, and p is *demonstrably not $BDL_{\rightarrow}$-derivable* from T, denoted $T \vdash_{BDL_{\rightarrow}} p^-$, if there is a proof tree for p^- in T using $BDL_{\rightarrow}$.

With the proof theory $BDL_{\rightarrow}$, the examples above can be correctly solved.

Example 13 Reconsider the defeasible theory from example 11. The proof tree below shows that $E\neg f$ is $BDL_{\rightarrow}$-derivable. A part of this proof tree deals with showing that $p \rightarrow \neg f$ is superior to $b \rightarrow f$. Because this example doesn't involve defeasible rules with empty antecedent, the third component in the labels can be omitted.



0.6.2 Ordered logic

The goal of ordered logic is to provide a theoretical foundation for knowledge based applications which support nonmonotonic or defeasible reasoning and which incorporate the knowledge of multiple experts in a principled way. The logic makes it possible to explicitly model internal perspectives or multiple agents and to resolve conflicts between competing perspectives without obscuring their opinions. While useful non-monotonic formalisms are available in which priorities are taken into account, they typically do not adress issues that arise when we want to incorporate the knowledge of several experts into our applications in a principled way. When we try to represent the knowledge of several experts in a single system, each expert has his own perspective on the relevant domain, and this difference in perspectives can lead to different conclusions. Even where knowledge of a single person is involved, a decision maker often has to take several conflicting perspectives into account when drawing conclusions on a certain body of evidence. One approach to representing multiple perspectives is to determine where conflicts arise and resolve them before building the knowledge representation. Another approach is to present the conclusions of all perspectives, leaving it to the user to make the final decisions. The best approach is a system that resolves differences and makes an overall recommendation in at least some cases, but that can also recover the viewpoints of the individual perspectives. In such a system, the conclusions drawn from a given perspective are defeasible and may be retracted when other perspectives are taken into account.

Examples of conflicting perspectives include conflicts between short- and longterm strategies or between strategies with different goals, such as situations where we might say, “As your teacher I must require you to hand in all assignments in this course, but as your friend I advise you to forget about the project for this class, take the lower grade, and concentrate on your other classes where you are in danger of failing.” Although it is possible to represent such conflicting perspectives as ordinary default rules in a single theory, this distorts the reality that there are really two different perspectives each of which leads to its own conclusions. We have the instructor’s perspective and the friendly advisor’s perspective. It may be helpful to derive the conclusions of each of the single perspectives even if there is not an overall conclusion that can be drawn in a particular case.

Ordered logic can be considered to be a proper extension of the implicit version of Nute’s basic defeasible logic and other formalisms based on implicit specificity information, by allowing a more complex precedence structure on rules. This precedence structure makes it possible to solve many examples of nonmonotonic reasoning [GVN94] for which the implicit formalisms fail to give an acceptable solution.

Although a credulous version of ordered logic exists [GV91], we will restrict the discussion here to the original approach to ordered logic, which is a skeptical one. Furthermore, we will concentrate on the proof-theoretical aspects of ordered logic. Most of the semantical aspects and a further elaboration of ordered logic are described in [Lae90, GLV91, LV90c, LSV90, LV90b, LV90a, Gee96]. In [GV93], a proof theory for nonmonotonic reasoning with implicit specificity information is defined following the ideas of ordered logic.

Ordered logic is defined for partially ordered defeasible theories. All rules are defeasible, and defeasible rules with empty antecents can be considered as observations. The partial order on nodes will be used to determine which if either of two defaults to apply in a theory when the two rules have contradictory consequents, by determining their precedence. $A \to p$ says that, other things being equal, we should accept p whenever we accept every member of A. Of course, it is the “other things being equal” that causes the problems. We could also say that in typical or normal cases where every member of A is true, so is p. But we can and often do adopt conflicting rules $A \rightarrow p$ and $B \rightarrow \neg p$, where it is possible that everything in both A and B is true. Such conflicts can only be resolved by giving one of the rules precedence over the other. One might interpret this as meaning that one rule is more reliable than the other, but that is not the interpretation we intend. Suppose $A \rightarrow p$ has precedence over $B \rightarrow \neg p$. This does not mean that $A \rightarrow p$ is more reliable than $B \rightarrow \neg p$ in the sense that we are better justified in adopting $A \rightarrow p$ than we are in adopting $B \rightarrow \neg p$. Each rule could be the very best possible rule for the case where its condition is satisfied, "all other things being equal". It is just when A and B are both satisfied, all things aren't equal where B is concerned. A situation where A and B are both satisfied may not be a typical or normal situation in which B is satisfied. The possibility of giving $A \rightarrow p$ precedence over $B \rightarrow \neg p$, by putting $A \rightarrow p$ at a strictly higher node, offers a way of solving an ambiguity in a theory where intuitively there should not be one. Obvious examples of such theories are taxonomic hierarchies in which subclasses don't answer the description which is typical for the class to which they belong, such as the penguin example (example 11). Whereas examples of this kind can also be correctly solved by specificity-based formalisms, the explicit priority structure can be used to solve many other problems.

Another way of looking at the partial order is as an "influence" relation between perspectives. If ω_i, ω_j and ω_k are perspectives in Ω with $\omega_i < \omega_k$ and $\omega_j < \omega_k$, then ω_k is a perspective that is influenced by perspectives ω_i and ω_j.

Typically, there will be a top perspective ω_0 such that $\omega_i \leq \omega_0$ for all perspectives $\omega_i \in \Omega$. ω_0 can be regarded as the final consolidation of all perspectives in the theory. Similarly, there may be a unique bottom node. Again our proof trees will need positive conclusions like "p holds" (denoted as p^+) and negative ones like "demonstrably, p does not hold" (denoted as p^-). Conclusions are derived with respect to a certain node, and since each node in an ordered theory can represent a distinct perspective, different conclusions will normally be derivable at different nodes. The final integrated conclusions are the ones that hold in the unique top node (if any). In contrast to Nute's approach, no evidentiality symbol is used.

Definition 17 Let $T = (\Omega, \leq, R, f)$ be a partially ordered defeasible theory, p a literal and $s \in \{+, -\}$. An OL-*proof tree* for p^s at a node ω_i in T is a finite tree [8] where each node is labeled q^t, where q is a literal and t is $+$ or $-$, such that the root is labeled p^s and each node m satisfies one of the following conditions:

[8] This proof theory is similar to a theory presented in [VNG89b], but there is an important difference. In the version presented here, rules at higher perspectives have precedence over rules at lower perspectives. In the earlier version, higher perspectives could only "see" lower perspectives, but lower perspectives took precedence. This is less natural than the current approach for modeling multiagent reasoning, but it is a promising theory for defeasible object oriented programming, see [LVVC89]. The proof theory presented here can be found in [GVN94], where the symbol $\Rightarrow$ is used for defeasible rules.

(OL1) m is labeled q^+ and $\exists A \to q \in f(\omega_j)$, where $\omega_j \leq \omega_i$ such that

1. m has a child node labeled a^+, for each $a \in A$; and
2. $\forall B \to \neg q \in f(\omega_k)$ where $\omega_k \leq \omega_i$ and $\omega_k \not< \omega_j$, $\exists b \in B$ such that m has a child node labeled b^-;

(OL2) m is labeled q^- and $\forall A \to q \in f(\omega_j)$, where $\omega_j \leq \omega_i$, either

1. m has a child node labeled a^- for some $a \in A$; or
2. $\exists B \to \neg q \in f(\omega_k)$, where $\omega_k \leq \omega_i$ and $\omega_k \not< \omega_j$, such that m has a child node labeled b^+ for each $b \in B$.

(OL3) m is labeled q^- and m has an ancestor labeled q^- such that there are no positively labeled nodes in between.

Condition (OL1) expresses defeasible rule application: a rule can be applied only if its antecedent holds and it is not defeated by an applicable competing rule. Condition (OL2) states that we can show that a literal doesn't hold if all rules that could conclude it are either not applicable or defeated by a competing rule. Condition (OL3) allows one to conclude p^-, when the only way to satisfy p is to satisfy p, which is e.g. the case for a theory containing a single rule $p \to p$.

Intuitively, the existence of a proof tree for p^+ at a perspective ω in T means that p is provable at ω in T. The existence of a proof tree for p^- at perspective ω in T means that we can show that p cannot be proven at ω in T.

Definition 18 Let $T = (\Omega, \leq, R, f)$ be a partially ordered defeasible theory. A literal p is *OL-derivable* from T at node ω, denoted $T \vdash_\omega p^+$ if there is an OL-proof tree for p^+ at ω. Such a literal p is also called an *OL-consequence* of T at ω. A literal p is *demonstrably not OL-derivable* from T at ω, denoted $T \vdash_\omega p^-$ if there is an OL-proof tree for p^- at ω. When the node under consideration is the unique top node, we simply say that p is OL-derivable from T (demonstrably not OL-derivable from T), denoted $T \vdash_{OL} p^+$ ($T \vdash_{OL} p^-$).

Example 14 Consider the ordered theory

$$(\{\omega_0, \omega_1, \omega_2\}, \leq, R, f)$$

where

$$\omega_2 < \omega_1 < \omega_0$$

and

$$\begin{aligned} f(\omega_0) &= \{\to s\} \\ f(\omega_1) &= \{\to b, s \to \neg r\} \\ f(\omega_2) &= \{b \to r\} \end{aligned}$$

In this example, r stands for rain, s for sunny and b for being in Belgium. This theory can be interpreted in different ways. From a single perspective point of view, a person could be walking in the borderland between Belgium and France, without knowing exactly on which side of the border he is. He knows that Belgium is a country where it frequently rains. However he believes that, when the sun is shining, it will not be raining, regardless whether he is in Belgium or not . At node ω_1, he assumes that he is in Belgium, but he knows nothing about the weather. Therefore, he concludes that it will probably be raining: the rule $b \rightarrow r$ is applicable at ω_1 while its competing rule $s \rightarrow \neg r$ is not . When suddenly the sun starts to shine brightly, he becomes very sure about the weather, information captured at node ω_0. Therefore, with the information available at node ω_0, he will conclude that it will not be raining: $s \rightarrow \neg r$ is applicable at ω_0 and all rules at or below ω_0 with consequent r are weaker (i.e. at a node below ω_1). The same conclusions can be made when we look at the example as containing the knowledge of two experts. Expert 1, who finds himself in a darkened room, has the knowledge contained in perspective ω_1 and concludes that it rains. Expert 2 can take a look outside and sees that the sun is shining. He knows more than expert 1, namely what is available at perspective ω_0, and concludes that it doesn't rain.

Example 15 Consider the ordered theory

$$(\{\omega_0, \omega_1, \omega_2\}, \leq, R, f)$$

where

$$\omega_1 < \omega_0, \omega_2 < \omega_0$$

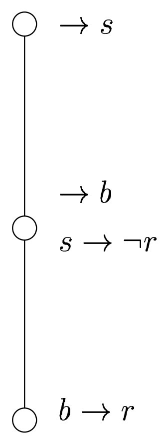
and

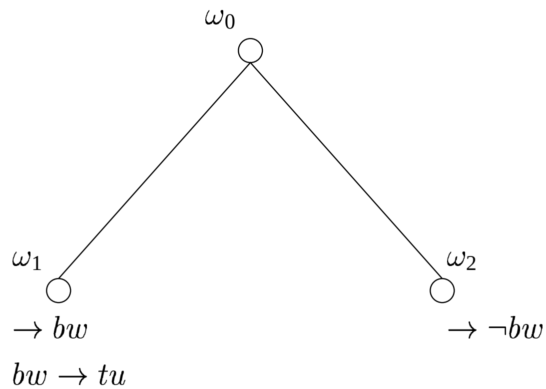
$$\begin{aligned} f(\omega_1) &= \{\rightarrow bw, bw \rightarrow tu\} \\ f(\omega_2) &= \{\rightarrow \neg bw\} \end{aligned}$$

This is a typical example of a perspective (ω_0) that is influenced by two other perspectives (ω_1 and ω_2). When asking advice about the weather prospects for tomorrow, the "expert" at perspective ω_1 believes that the weather will be bad (bw) and that you should take an umbrella (tu) with you when you go for a walk. The "expert" at ω_2 believes the weather won't be bad. Therefore, at perspective ω_1, the conclusion tu holds where tu is obtained by applying $bw \rightarrow tu$. At perspective ω_0, the condition bw used to derive tu at ω_1 does not hold since it is defeated by $\rightarrow \neg bw$ at ω_2. Rule $bw \rightarrow tu$ is therefore not applicable at ω_0, and tu cannot be proven at ω_0.

This proof theory is well-behaved, i.e. we can show that no literal is at the same time $OL-$derivable and demonstrably not $OL-$derivable. In other words, when p is a literal, we don't have $T \vdash_\omega p^+$ and $T \vdash_\omega p^-$ at the same time.

However, it can be the case that nothing can be proven about some literals, as shown in the following example.





Example 16 Consider the ordered theory

$$(\{\omega_0\}, \emptyset, \{\rightarrow p, p \rightarrow \neg p\}, f)$$

with

$$f(\omega_0) \;=\; \{\rightarrow p, p \rightarrow \neg p\}$$

It turns out that p doesn't hold at ω_0, because there is no OL-proof tree of p^+ at ω_0. However, using OL, we cannot show that p doesn't hold, because there is no OL-proof tree of p^- at ω_0.

Furthermore, it can be shown that two complementary literals cannot be $OL-$derivable at the same time. In other words, when p is a literal, we don't have both $T \vdash_\omega p^+$ and $T \vdash_\omega \neg p^+$.

0.6.3 Explicit version of Nute's basic defeasible logic

In his family of defeasible logics, Nute also surveys a logic for dealing with explicit priorities [Nut92, GVN94], which we will call $EBDL$. An important difference with ordered logic is that in $EBDL$, the partial order is given on the set of defeasible rules and defeaters, instead of on nodes [9]. When a defeasible rule competes with a strict rule, the defeasible rule is always defeated. To adjudicate between two defeasible rules or between a defeasible rule and a defeater, the partial order is consulted.

Definition 19 Let $T = (O, R)$ be a defeasible theory and let $\leq$ be a partial order on $R_d \cup R_i$. Where p is a sentence (i.e. a literal or an E-sentence) and s is $+$ or $-$, a *proof tree* for p^s in T using $EBDL$ is a finite tree where each node is labeled q^t, where q is a sentence and t is $+$ or $-$, such that the root is labeled p^s and each node m satisfies one of the following conditions:

[9] In [Nut92] the logic $EBDL$ is originally called $SD_<$, standing for the set of conditions $\{M^+, M^-, E^+, E^-, SS^+, D^+_<, SD^-_<\}$.

(D1) m is labeled q^+ and either $q \in O$ or there is a strict rule $A \Rightarrow q \in R_s$ such that for each $a \in A, m$ has a child node labeled a^+.

(D2) m is labeled $q^-, q \notin O$, and for every strict rule $A \Rightarrow q \in R_s$, there is $a \in A$ and a child node of m labeled a^-.

(D3) m is labeled Eq^+ and m has a child node labeled q^+.

(D4) m is labeled Eq^+ and m has a child node labeled $\neg q^-$ and there is a strict rule $A \Rightarrow q \in R_s$ such that for each $a \in A, m$ has a child node labeled Ea^+ and for each $B \Rightarrow \neg q \in R_s$, there is $b \in B$ and a child node of m labeled Eb^-.

(D5) m is labeled Eq^+ and m has a child node labeled $\neg q^-$ and there is a defeasible rule $A \rightarrow q \in R_d$ such that

1) for each $a \in A, m$ has a child node labeled Ea^+, and
2) for each rule $r \in R_s \cup R_d \cup R_i$ with $H(r) = \neg q$, either $r < A \rightarrow q$, or there is $c \in B(r)$ and a child node of m labeled Ec^-.

(D6) m is labeled Eq^-, m has a child node labeled q^- and m has a child node labeled $\neg q^+$.

(D7) m is labeled Eq^-

1) m has a child node labeled q^- and
2) for each rule $r \in R_s \cup R_d$ with $H(r) = q$, either
   a) there is $a \in B(r)$ and a child node of m labeled Ea^-, or
   b) there is a rule $r' \in R_s \cup R_d \cup R_i$ such that $H(r') = \neg q, r' \not< r$ and m has a child node labeled Eb^+ for each $b \in B(r')$.

This proof theory is similar to the one for $BDL_{\rightarrow}$ using defeasible specificity, except for conditions (D5) and (D7), where now the explicit partial order is used for resolving conflicts instead of specificity-based arguments. This explicit partial order is also the reason why we no longer need to include a set of literals into the labels. Note that the partial order is given on defeasible and interfering rules, so that no defeasible rule is ever superior to any strict rule in conditions (D5) and (D7).

Definition 20 Where $T = (O, R)$ is a defeasible theory, $\leq$ a partial order on $R_d \cup R_i$ and p a sentence (i.e. a literal or an E-sentence), p is *$EBDL$-derivable* from T using $\leq$, denoted $T \vdash_{EBDL} p^+$, if there is a proof tree for p^+ in T using $EBDL$ and $\leq$, and p is *demonstrably not $EBDL$-derivable* from T using $\leq$, denoted $T \vdash_{EBDL} p^-$, if there is a proof tree for p^- in T using $EBDL$ and $\leq$.

0.6.4 The preemption problem

Ordered logic and all versions of basic defeasible logic presented thus far all suffer from the preemption problem. This problem emerges in the context of inheritance reasoning with exceptions. Let us illustrate the notion of preemption by means of an example [GVN94]. Suppose that normally a computer science professor at a junior college is poor, even though computer science professors at junior colleges normally have a Ph.D. in computer science, and people who have a Ph.D. in computer science normally are not poor. Suppose also that ne'er-do-well, disinherited scions of wealthy families normally are poor, even though such individuals are clearly scions of wealthy families, and scions of wealthy families normally are not poor. If John is both a computer science professor at a junior college and a ne'er-do-well, disinherited scion of a wealthy family, we would intuitively conclude that John is poor. However, if we represent this knowledge in an inheritance network, we have to face the problem [THT87] that none of the positive links to the property "poor" is more specific than both negative links to the same property. The same problem arises in ordered logic and Nute's defeasible logic, after translating this inheritance network into an ordered theory, and adapting a skeptical attitude. To solve this problem, inheritance reasoners like SIR [HTT87] impose the requirement that a compound path is permitted only when every conflicting path is preempted. More specific, SIR requires that a positive path is permitted provided only that the part up to the last link is permitted and for every negative link competing with the last link, either there is no permitted positive path up to its start node or there is another positive link such that there is a permitted path through the start node of this positive link ending in the start node of the negative link. Using this principle, we arrive at the conclusion that John is poor in the inheritance network resulting from the example.

We can translate this principle into ordered logic by making the restriction that a rule can only be defeated by a competing rule which is not itself defeated by a strict competitor. This preemption principle can easily be integrated into the skeptical proof theory for ordered logic, yielding the following proof theory.

Definition 21 Let $T = (\Omega, \leq, R, f)$ be a partially ordered defeasible theory, p a literal and $s+$ or $-$. A proof tree for p^s at node ω_i in T is a finite tree where each node is labeled q^+, where q is a literal and t is $+$ or $-$, such that the root is labeled p^s and each node m sastisfies one of the following conditions:

(OL1) m is labeled q^+ and $\exists A \rightarrow q \in f(\omega_j)$, where $\omega_j \leq \omega_i$ such that

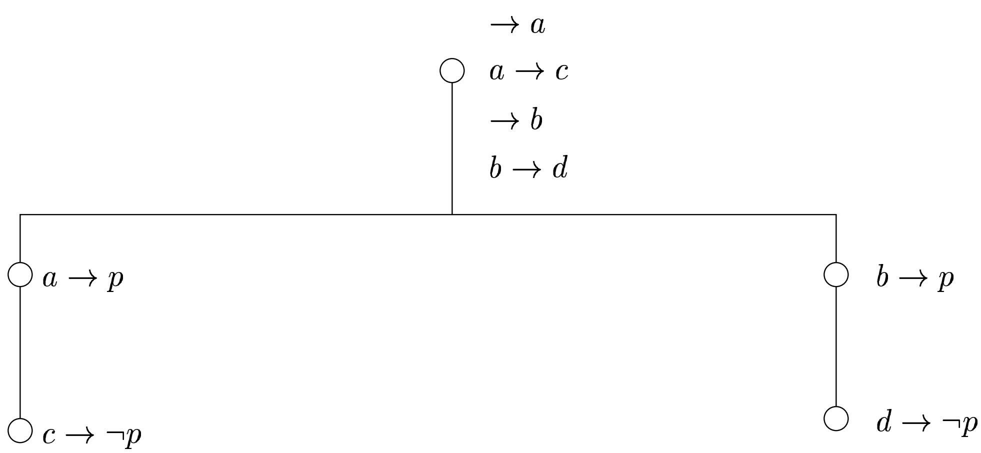
1. m has a child node labeled a^+, for each $a \in A$; and
2. $\forall B \rightarrow \neg q \in f(\omega_k)$ where $\omega_k \leq \omega_i$ and $\omega_k \not< \omega_j$, $\exists b \in B$ such that m has a child node labeled b^-, or there is a rule $C \rightarrow q \in f(\omega_l)$, where $\omega_l \leq \omega_i$ and $\omega_l > \omega_k$, such that m has a child node labeled c^+ for each $c \in C$.

(OL2) m is labeled q^- and for each rule $A \rightarrow q \in f(\omega_j)$, where $\omega_j \leq \omega_i$, either

1. m has a child node labeled a^- for some $a \in A$; or
2. $\exists B \rightarrow \neg q \in f(\omega_k)$, where $\omega_k \leq \omega_i$ and $\omega_k \not< \omega_j$, such that m has a child node labeled b^+ for each $b \in B$, and for each rule $C \rightarrow p \in f(\omega_l)$ where $\omega_l \leq \omega_i$ and $\omega_l > \omega_k, m$ has a child node labeled c^- for some $c \in C$.

(OL3) m is labeled q^- and m has an ancestor labeled q^- such that there are no positively labeled nodes in between.

If we introduce the propositions p (poor), a (computer science professors at junior colleges), b (disinherited ne'er-do-well scions of wealthy families), c (holders of a Ph.D. in computer science) and d (scions of wealthy families), we can formalize the example in the following ordered theory. The original OL proof theory, given in definition 17,



would derive no conclusion about whether John is poor or not, while the proof theory extended with the preemption principle arrives at the intuitively correct conclusion that John is poor.

Similar solutions are presented for the family of basic defeasible logics [Nut92].

0.6.5 Ryan's formalism of ordered theories presentations

Ryan [Rya92] proposes a framework for reasoning with ordered theory representations which is comparable to ordered logic. Instead of providing an ordering on sets of rules, the ordering is given on sentences. Whereas ordered logic is originally a directly skeptical formalism, Ryan's formalism can be considered to be indirectly skeptical: the entailed conclusions are the ones that hold in all maximal models, where maximality is understood according to an ordering of interpretations which favours those satisfying as many (high priority) sentences as possible. Due to the use of sentences instead of sets of rules, another ordering is required: for each sentence, a satisfaction ordering on interpretations is defined reflecting the degree to which an interpretation satisfies the sentence.

0.7 Summary and discussion

The defeasible logics discussed in this chapter can be grouped into two families, according to how they try to eliminate ambiguities. One family, containing Pearl's system Z, Geffner's conditional entailment, Simari's and Loui's argument-based system and Nute's basic defeasible logics $BDL_{\rightarrow}$ and $BDL_{\Rightarrow}$, relies on specificity information implicitly present in the knowledge base in order to solve conflicts. This family can be summarized in the following table, in which we emphasize the different basic design choices.

IMPLICIT PRIORITIES				
	Z	CE	ABS	$BDL_{\rightarrow}$
subject of priority	sets of defaults	defaults	argument structures	conflicting defaults
kind of priority	total	irreflexive+ transitive	irreflexive+ transitive	irreflexive+ transitive
attitude	skeptical	skeptical (indirect)	skeptical amb.prop.	skeptical amb.block.
contraposition	Y	Y	N	N
observations	Y	Y	Y	Y
strict rules	N	Y	Y	Y
default rules	Y	Y	Y	Y
kind of specificity	defeasible	defeasible	$\pm$strict	defeasible

The second family, containing Brewka's system of preferred subtheories, Nute's basic defeasible logic EBDL, ordered logic and Ryan's formalism of ordered theory presentations, makes use of an additional structure to represent explicit priorities.

EXPLICIT PRIORITIES			
	PS	OL	EBDL
subject of priority	sets of defaults	sets of defaults	defaults+ defeaters
kind of priority	total/ partial	partial	irreflexive+ transitive
attitude	credulous	skeptical/ credulous	skeptical
contraposition	Y	N	N

Both uses of prioritization have their pros and cons. An approach in which priorities are explicitly supplied by the user can be useful, because the flexible means for deciding among competing defaults allows us to give solutions for several examples of nonmonotonic reasoning which cannot be solved using implicit specificity information. Explicit priorities are useful when preference criteria other than specificity, such as recency, authority, reliability, .. are required. However, when specificity is the preference criterion, the user finds himself obliged to perform the redundant task of explicitly providing priority information which is implicitly present in the knowledge base. This explicitly given priority information might even contradict the implicitly present priorities, which may not always be as intended. Therefore, both approaches need to be considered.

Recently, several attempts have been made to combine the best of both worlds into a single formalism. System Z^+ [GP91] relies on implicit specificity information but allows explicit priorities in the sense of additional rule-strenghts. Here, strenghts can be used to refine the specificity-based priorities. They undergo adjustments, so that compliance with specificity-type constraints is automatically preserved. As a result, specificity can never be overridden, or defeated. However, it has been argued that sometimes it might be necessary to give precedence to selection criteria other than specificity. E.g. in the area of legal reasoning [Bre94], it might be the case that a more recent general law overrides a more specific older law, i.e. that the recency criterion is stronger than the specificity criterion. A formalism able to deal with this kind of reasoning is presented in [GV95]. In this argument-based formalism, specificity is considered to be the preference criterion by default. However, additional priorities can be added from the very beginning, making it possible to refine or even defeat the specificity criterion. Furthermore, this formalism allows to reason about priorities. By presenting priorities within the logical language [Bre94], statements concerning priorities can be derived, but also defeated.

Bibliography

[Ada75] E. Adams. *The logic of conditionals*. 1975.

[BCD+93] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax-based entailment. In *Proceedings 13th International Joint Conference on Artificial Intelligence*, pages 640–645, 1993.

[BDP93] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Ninth annual conference on Uncertainty in Artificial Intelligence (UAI)*, pages 411–419, 1993.

[Bre89] G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proceedings IJCAI-89*, 1989.

[Bre94] Gerhard Brewka. Reasoning about priorities in default logic. Technical report, GMD, Postfach 13 16, 53731 Sankt Augustin, Germany, 1994.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Gee96] P. Geerts. *Ordered logic and prioritization in nonmonotonic reasoning*. PhD Thesis. University of Antwerp, UIA, 1996.

[Gef92] H. Geffner. *Default Reasoning: Causal and Conditional Theories*. MIT Press, 1992.

[Gin94] M.L. Ginsberg. *AI and Nonmonotonic Reasoning*, volume Handbook of Logic in Artificial Intelligence and Logic Programming, pages 1–33. Oxford University Press, 1994.

[GLV91] D. Gabbay, E. Laenens, and D. Vermeir. Credulous vs. sceptical semantics for ordered logic programs. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 208–217, Cambridge, Mass, 1991. Morgan Kaufmann.

[GP91] M. Goldszmidt and J. Pearl. System z^+: a formalism for reasoning with variable-strength. In *Proceedings AAAI-91*, pages 399–404, 1991.

[GP92] H. Geffner and J. Pearl. Conditional entailment: bridging two approaches to default reasoning. *Artificial Intelligence*, 53:209–244, 1992.

[GV91] P. Geerts and D. Vermeir. Credulous and autoepistemic reasoning using ordered logic. In *Proceedings of the First International Workshop on Logic Programming and Non-Monotonic Reasoning*, pages 21–36. MIT Press, 1991.

[GV93] P. Geerts and D. Vermeir. A nonmonotonic reasoning formalism using implicit specificity information. In *Proceedings of the second International Workshop on Logic Programming and Non-monotonic Reasoning*. The MIT Press, 1993.

[GV95] Patricia Geerts and Dirk Vermeir. Specificity by default. In C. Froidevaux and J. Kuhlas, editors, *Symbolic and Quantitative Approaches to Reasoning and Uncertainty, ECSQARU95*, number 946 in Lecture notes in Artificial Intelligence, pages 107–216. Springer Verlag, July 1995.

[GVN94] P. Geerts, D. Vermeir, and D. Nute. Ordered logic: defeasible reasoning for multiple agents. *Decision Support Systems*, 11:157–190, 1994.

[HTT87] J.F. Horty, R.H. Thomason, and D.S. Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. Technical Report CMU report CMU-CS-87-175, 1987.

[KLM90] S. Kraus, D. Lehman, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Lae90] E. Laenens. Foundations of ordered logic, 1990. PhD Thesis, University of Antwerp UIA.

[Lif85] V. Lifschitz. Computing circumscription. In *Proceedings 9th IJCAI*, pages 121–127. Los Angeles, CA, 1985. Also in 'Readings in nonmonotonic reasoning', M.L. Ginsberg.

[LM88] D. Lehmann and M. Magidor. Rational logics and their models: a study in cumulative logics. Technical Report TR-88-16, Dept. of Computer Science, Hebrew University, Jerusalem, Israel, 1988.

[Lou87] R. Loui. Defeat among arguments: A system of defeasible inference. *Computational Intelligence*, 3(2):100–106, 1987.

[LSV90] E. Laenens, D. Sacca, and D. Vermeir. Extending logic programming. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 184–193. ACM, Association for Computing Machinery, 1990.

[LV90a] E. Laenens and D. Vermeir. Assumption-free semantics for ordered logic programs: on the relationship between well-founded and stable partial models, 1990. University of Antwerp, Tech. Report 90-19.

[LV90b] E. Laenens and D. Vermeir. A fixpoint semantics of ordered logic. *Journal of Logic and Computation*, 1(2):159–185, 1990.

[LV90c] E. Laenens and D. Vermeir. A logical basis for object-oriented programming. In *Proc. European Workshop on Logics in AI, Amsterdam*, pages 317–332. Springer Verlag, 1990.

[LVVC89] E. Laenens, B. Verdonk, D. Vermeir, and A. Cuyt. A logic for objects and inheritance. In *Proc. of the Advanced Database Symposium, Kyoto*, pages 55–60. Information Processing Society of Japan, 1989.

[Mak89] D. Makinson. General theory of cumulative inference. In *Proceedings 2nd International Workshop on nonmonotonic reasoning*, pages 1–18. Springer Verlag, 1989.

[McC80] J. McCarthy. Circumscription - a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980.

[McC86] J. McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28:89–116, 1986. Also in 'Readings in nonmonotonic reasoning', M.L. Ginsberg.

[MD80] D. McDermott and J. Doyle. Non-monotonic logic i. *Artificial Intelligence*, 13:41–72, 1980.

[Moo84] R.C. Moore. Possible-world semantics for autoepistemic logic. In *Proceedings 1984 Non-monotonic reasoning workshop (AAAI)*, pages 344–354. new Paltz, NY, 1984. Also in 'Readings in nonmonotonic reasoning', M.L. Ginsberg.

[Moo88] R.C. Moore. Autoepistemic logic. In P. Smets, editor, *Non-standard logics for automated reasoning*, pages 105–136. 1988.

[Nut85] D. Nute. Conditional logic. In D. Gabbay and F. Guenthner, editors, *Handbook of philosophical logic II*. 1985.

[Nut86] D. Nute. LDR: a logic for defeasible reasoning, 1986. ACMC Research Report 01-0013.

[Nut88] D. Nute. Defeasible reasoning and decision support systems. *Decision support systems*, 4:97–110, 1988.

[Nut90] D. Nute. *Defeasible logic and the frame problem*, volume Knowledge representation and defeasible reasoning, pages 3–21. Kluwer academic publishers, 1990.

[Nut92] D. Nute. *Basic defeasible logic*, pages 125–154. Intensional Logics for Logic Programming. Oxford University Press, 1992.

[Pea88] J. Pearl. *Probabilistic reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, 1988.

[Pea89] J. Pearl. Probabilistic semantics for nonmonotonic reasoning : A survey. In *Proceedings 1st International Conference on Principles of Knowledge Representation and Reasoning*. Toronto, 1989.

[Pea90] J. Pearl. System z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Theoretical Aspects of Reasoning about Knowledge (TARK-III)*, pages 121–135. Morgan Kaufmann, 1990.

[Pol87a] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[Pol87b] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[Pol92] J.L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[Poo85] D. Poole. On the comparison of theories: Preferring the most specific explanation. In *Proceedings IJCAI-85*. 1985.

[Poo88] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[Rei80] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Res64] N. Rescher. *Hypothetical reasoning*. North-Holland Publ., Amsterdam, 1964.

[Rya92] M. Ryan. Ordered presentations of theories, 1992. PhD Thesis.

[Sho87a] Y. Shoham. *Reasoning about change*. MIT Press, 1987.

[Sho87b] Y. Shoham. A semantical approach to nonmonotonic logics. In *Proceedings of the Symposium on Logic in Computer Science*, pages 275–279. 1987.

[SL92] Guillermo R. Simari and Ronald P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53:125–157, 1992.

[Ste92] L.A. Stein. Resolving ambiguity in nonmonotonic inheritance hierarchies. *Artificial Intelligence*, 55:259–310, 1992.

[THT87] D.S. Touretzky, J.F. Horty, and R.H. Thomason. A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. In *Proceedings IJCAI-87*. 1987.

[VNG89a] D. Vermeir, D. Nute, and P. Geerts. A defeasible logic for multi-expert systems. In *Proc. of the Intl. Symposium on Computational Intelligence 89*. Elsevier Publ. Co., 1989.

[VNG89b] D. Vermeir, D. Nute, and P. Geerts. A logic for defeasible perspectives. In *Proc. of the 1988 Tubingen Workshop on Semantic Networks and Nonmonotonic Reasoning, Vol. 1*, pages 1–27. SNS-Bericht 89-48, 1989.

[VNG90] D. Vermeir, D. Nute, and P. Geerts. Modeling defeasible reasoning with multiple agents. In *Proc. of the HICSS, Vol. III*, pages 534–543. 1990.

[Vre91] G. Vreeswijk. The feasibility of defeat in defeasible reasoning. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR91)*, pages 526–534, 1991.