A General Approach to Specificity in Default Reasoning

James P. Delgrande
School of Computing Science
Simon Fraser University
Burnaby, B.C.
Canada V5A 1S6
jim@cs.sfu.ca

Torsten H. Schaub
IRISA
Campus de Beaulieu
35042 Rennes cedex
France
torsten@irisa.fr

Abstract

We present an approach addressing the notion of specificity, or of preferring a more specific default sentence over a less specific one, in commonsense reasoning. Historically, approaches have either been too weak to provide a full account of defeasible reasoning while accounting for specificity, or else have been too strong and fail to enforce specificity. Our approach is to use the techniques of a weak system, as exemplified by System **Z**, to isolate minimal sets of conflicting defaults. From the specificity information intrinsic in these sets, a default theory in a target language is specified. In this paper we primarily deal with theories expressed (ultimately) in Default Logic. However other approaches would do just as well, as we illustrate by also considering Autoepistemic Logic and variants of Default Logic. In our approach, the problems of weak systems, such as lack of adequate property inheritance and (occasional) unwanted specificity relations, are avoided. Also, difficulties inherent with stronger systems, in particular, lack of specificity are addressed. This work differs from previous work in specifying priorities in Default Logic, in that we obtain a theory expressed in Default Logic, rather than ordered sets of rules requiring a modification to Default Logic.

1 Introduction

A general problem in nonmonotonic reasoning is that *specificity* among default assertions is difficult to obtain in a fully satisfactory manner. Consider for example where birds fly, birds have wings, penguins are birds, and penguins don't fly. We can write this as:

$$B \rightarrow F,\ B \rightarrow W,\ P \rightarrow B,\ P \rightarrow \neg F. \qquad (1)$$

From this theory, given that P is true, one would want to conclude $\neg F$ by default. Intuitively, being a penguin is a more specific notion than that of being a bird, and, in the case of a conflict, we would want to use the more specific default. Also, given that P is true one would want to conclude that W was true, and so penguins have wings by virtue of being birds.

Autoepistemic Logic [Moore, 1985], Circumscription [McCarthy, 1980], and Default Logic [Reiter, 1980] are examples of approaches that are overly *permissive*. For example, in the obvious representation of the above theory in Default Logic, we obtain one extension (i.e. a set of default conclusions) in which $\neg F$ is true and another one in which F is true. One is required to use so-called semi-normal defaults[1] to eliminate the second extension. [Reiter and Criscuolo, 1981], for example, gives a list of ways of transforming default theories so that unwanted extensions arising from specific "interactions" are eliminated.

In the past few years there has been some consensus as to what should constitute a basic system of default properties. This, arguably, is illustrated by the convergence (or at least similarity among) systems such as those developed in [Delgrande, 1987; Kraus *et al.*, 1990; Pearl, 1990; Boutilier, 1992a; Geffner and Pearl, 1992], yet which are derived according to seemingly disparate intuitions. A general problem with these accounts however is that they are too weak. Thus in a conditional logic, even though a bird may be assumed to fly by default (i.e. in the preceding theory, we only derive F but not $\neg F$), a green bird cannot be assumed to fly by default (since it is *conceivable* that greenness is relevant to flight). In these systems some mechanism is required to assert that properties not known to be relevant are irrelevant. This is done in conditional logics by meta-theoretic assumptions, and in probabilistic accounts by independence assumptions. In other approaches there are problems concerning property inheritance, and so one may not obtain the inference that a penguin has wings. While various solutions have been proposed, none are entirely satisfactory.

[1] See Section 2.2 for a definition of semi-normal defaults and the way they deal with unwanted extensions.

Our approach is to use the specificity information given by a "weak" system to generate a default theory in a "strong" system, where specificity and property inheritance are satisfactorily handled. Hence we address two related but essentially independent questions:

1. How can a (so-called) weak system be used to isolate specific interacting defaults?
2. How can this information be uniformly incorporated in a theory expressed in a (so-called) strong system?

For concreteness, we develop the approach by considering System **Z** [Pearl, 1990] as an example of a weak system of defeasible reasoning, and Default Logic (**DL**) [Reiter, 1980] as a strong system; however in Section 6 we consider the application of the approach to other systems. The general idea is to combine the techniques of System **Z** and **DL** in a principled fashion to obtain a general hybrid approach for defeasible reasoning.

We begin with a set of default conditionals $R = \{r \mid \alpha_r \rightarrow \beta_r\}$ where each α_r and β_r are arbitrary propositional formulas. By means of System **Z** we isolate minimally conflicting sets of defaults with differing specificities; intuitively the defaults in such a set should never be simultaneously applicable. Notably we do not use the full ordering given by System **Z** (which has difficulties of its own, as described in the next section), but rather appeal to the *techniques* of this approach to isolate conflicting subsets of the defaults. In a second step, we use the derived specificity information to produce a set of semi-normal default rules in **DL** from the rules in R, in such a way that specificity is suitably handled. The framework described here is intended to be a general approach to "compiling" default theories expressed by a set of conditionals, using intuitions from a weak approach (exemplified by System **Z**), into a strong approach (exemplified by **DL**). The choice of **DL** is of course not arbitrary, since it is very well studied and there exist implementations of **DL**.

The specific approach then can be looked at from two perspectives. First, **DL** is used to circumvent problems in System **Z**, including the facts that inheritance isn't possible across conflicting subclasses and that unwanted specificity information may be obtained. Second, System **Z** is used to address problems in **DL** that arise from interacting defaults. That is, using System **Z**, we construct theories in **DL** wherein specificity is appropriately handled. Hence, this paper might in some respects be looked on as a successor to [Reiter and Criscuolo, 1981], in that the situations addressed here subsume the set of modifications suggested in that paper. Moreover, the present approach provides a justification for these modifications.

Specificity information is thus obtained by appeal to an extant theory of defaults (here, System **Z**), and not some a priori ordering. In addition, and in contrast to previous approaches, specificity is added to **DL** without changing the machinery of **DL**. That is, the resultant default theory is a theory *in* **DL**, and not a set of ordered default rules requiring modifications to **DL**. Finally, we do not produce a "global" partial order (or orders) of rules but rather "locally" distinguish conflicting rules. Lastly, specificity conflicts are resolved, leaving unchanged other conflicts (as are found for example in a "Nixon diamond").

In the next section we briefly introduce System **Z**, Default Logic, and related work. Section 3 introduces and develops our approach, while Sections 4 and 5 provide the formal details. Section 6 considers the application of the approach to other systems. Section 7 gives a brief summary.

2 Background

2.1 System Z

In System **Z** a set of rules R representing default conditionals is partitioned into an ordered list of mutually exclusive sets of rules $R_0, \ldots, R_n$. Lower ranked rules are considered more normal (or less specific) than higher ranked rules. Rules appearing in lower-ranked sets are *compatible* with those appearing in higher-ranked sets, whereas rules appearing in higher-ranked sets *conflict* in some fashion with rules appearing in lower-ranked sets. One begins then with a set $R = \{r \mid \alpha_r \rightarrow \beta_r\}$ where each α_r and β_r are propositional formulas over a finite alphabet.[2] A set $R' \subseteq R$ *tolerates* a rule r if $\{\alpha_r \wedge \beta_r\} \cup R'$ is satisfiable. We assume in what follows that R is **Z**-*consistent*,[3] i.e. for every non-empty $R' \subseteq R$, some $r' \in R'$ is tolerated by $R' - \{r'\}$. Using this notion of tolerance, an ordering on the rules in R is defined:

1. First, find all rules tolerated by R, and call this subset R_0.
2. Next, find all rules tolerated by $R - R_0$, and call this subset R_1.
3. Continue in this fashion until all rules have been accounted for.

In this way, we obtain a *partition* $(R_0, \ldots, R_n)$ of R, where $R_i = \{r \mid r \text{ is tolerated by } R - R_0 - \ldots - R_{i-1}\}$ for $1 \leq i \leq n$. More generally, we write R_i to denote the ith set of rules in the partition of a set of conditionals R. A set of rules R is called *trivial* iff its partition consists only of a single set of rules.

The rank of rule r, written $\mathbf{Z}(r)$, is given by: $\mathbf{Z}(r) = i$ iff $r \in R_i$. Every model M of R is given a **Z**-rank, $\mathbf{Z}(M)$, according to the highest ranked rule it falsifies:

$$\mathbf{Z}(M) = min\{\ n \mid M \models \alpha_r \supset \beta_r,\ \mathbf{Z}(r) \geq n\}.$$

[2] The inclusion of strict rules is straightforward [Delgrande and Schaub, 1993] but for simplicity is omitted here.

[3] Pearl uses the term *consistent* [Pearl, 1990].

For our initial set of rules in (1), we obtain the ordering

$$R_0 = \{B \rightarrow F, B \rightarrow W\}, \quad (2)$$
$$R_1 = \{P \rightarrow B, P \rightarrow \neg F\}. \quad (3)$$

So the **Z** rank of the model in which B, $\neg F$, W, and P are true is 1, since the rule $B \rightarrow F$ is falsified. The **Z** rank of the model in which B, F, W, and P are true is 2, since the rule $P \rightarrow \neg F$ is falsified.

The rank of an arbitrary formula φ is defined as the lowest **Z**-rank of all models satisfying φ: $\mathbf{Z}(\varphi) = min\{\mathbf{Z}(M) \mid M \models \varphi\}$. Finally we can define a form of default entailment, which is called *1-entailment*, as follows: A formula φ is said to 1-entail ϕ in the context R, written $\varphi \vdash_1 \phi$, iff $\mathbf{Z}(\varphi \wedge \phi) < \mathbf{Z}(\varphi \wedge \neg\phi)$.

This gives a form of default inference that is weaker than Default Logic, yet has some very nice properties. In the preceding example, we obtain that $P \vdash_1 \neg F$, and $P \vdash_1 B$ and so penguins don't fly, but are birds. Unlike **DL**, we cannot infer that penguins fly, i.e. $P \not\vdash_1 F$. Irrelevant facts are also handled well (unlike conditional logics), and for example we have $B \wedge G \vdash_1 F$, so green birds fly. There are two weaknesses with this approach. First, one cannot inherit properties across exceptional subclasses. So one cannot conclude that penguins have wings (even though penguins are birds and birds have wings), i.e. $P \not\vdash_1 W$. Second, undesirable specificities are sometimes obtained. For example, consider where we add to our initial example (1) the default that large animals are calm. We get the **Z**-ordering:

$$R_0 = \{B \rightarrow F,\ B \rightarrow W,\ L \rightarrow C\}, \quad (4)$$
$$R_1 = \{P \rightarrow B,\ P \rightarrow \neg F\}. \quad (5)$$

Intuitively $L \rightarrow C$ is irrelevant to the other defaults, yet one obtains the default conclusion that penguins aren't large, since $\mathbf{Z}(L \wedge \neg P) < \mathbf{Z}(L \wedge P)$.

[Goldszmidt and Pearl, 1990] has shown that 1-entailment is equivalent to *rational closure* [Kraus *et al.*, 1990]; [Boutilier, 1992a] has shown that CO^* is equivalent to 1-entailment and that N [Delgrande, 1987] and $CT4$ are equivalent to the more basic notion of 0-entailment, proposed in [Pearl, 1989] as a "conservative core" for default reasoning. Consequently, given this "locus" of closely-related systems, each based on distinct semantic intuitions, these systems (of which we have chosen System **Z** as exemplar) would seem to agree on a principled minimal approach to defaults.

2.2 Default Logic

In Default Logic, classical logic is augmented by *default rules* of the form $\frac{\alpha : \beta}{\omega}$. Even though almost all "naturally occurring" default rules are normal, i.e. of the form $\frac{\alpha : \beta}{\beta}$, semi-normal default rules, of the form $\frac{\alpha : \beta \wedge \omega}{\beta}$, are required for establishing precedence in the case of "interacting" defaults [Reiter and Criscuolo, 1981] (see below). Default rules induce one or more *extensions* of an initial set of facts. Given a set of facts W and a set of default rules D, any such extension E is a deductively closed set of formulas containing W such that, for any $\frac{\alpha : \beta}{\omega} \in D$, if $\alpha \in E$ and $\neg\beta \notin E$ then $\omega \in E$.

Definition 1 *Let (D, W) be a default theory and let E be a set of formulas. Define $E_0 = W$ and for $i \geq 0$*

$$E_{i+1} = Th(E_i) \cup \left\{ \omega \;\middle|\; \frac{\alpha : \beta}{\omega} \in D, \alpha \in E_i, \neg\beta \notin E \right\}.$$

Then E is an extension for (D, W) iff $E = \bigcup_{i=0}^{\infty} E_i$.

The above procedure is not strictly iterative since E appears in the specification of E_{i+1}.

Consider our birds example (1); in **DL**, it can be expressed as:[4] $\frac{B : F}{F}$, $\frac{B : W}{W}$, $\frac{P : B}{B}$, $\frac{P : \neg F}{\neg F}$. Given that P is true, we obtain two extensions: one in which P, B, W, and F are true and another one in which P, B, W, and $\neg F$ are true. Intuitively we want only the last extension, since the more specific default $\frac{P : \neg F}{\neg F}$ should take precedence over the less specific default $\frac{B : F}{F}$. The usual fix is to establish a precedence among these two interacting defaults by adding the exception P to the justification of the less specific default rule. This amounts to replacing $\frac{B : F}{F}$ by $\frac{B : F \wedge \neg P}{F}$ which then yields the desired result, namely a single extension containing P, B, W, and $\neg F$.

2.3 Related Work

Arguably specificity per se was first specifically addressed in default reasoning in [Poole, 1985], although it has of course appeared earlier. Of the so-called "weak" approaches, as mentioned, we could have as easily used approaches described in [Boutilier, 1992a] or [Kraus *et al.*, 1990] as that of System **Z**; however specificity, as it appears in System **Z** is particularly straightforwardly describable. Other approaches are too weak to be useful here. For example conditional entailment [Geffner and Pearl, 1992] does not support full inheritance reasoning; while [Delgrande, 1988] is unsatisfactory since it gives a syntactic, albeit general, approach in the framework of conditional logics.

In Default Logic, [Reiter and Criscuolo, 1981] considers patterns of specificity in interacting defaults, and describes how specificity may be obtained via appropriate semi-normal defaults. This work in fact may be regarded as a pre-theoretic forerunner to the present approach, since the situations addressed therein all constitute instances of what we call (in the next section) minimal conflicting sets. [Etherington and Reiter, 1983] also considers a problem that fits within the (overall) present framework: specificity information is given by an inheritance network; this network is compiled into a default theory in **DL**.

[4] For coherence, we avoid strict implications which might be more appropriate for some of the rules.

Of recent work that develops priority orderings on default theories, we focus on the approaches of [Boutilier, 1992b; Baader and Hollunder, 1993a; Brewka, 1993]. We note however that these approaches obtain specificity by requiring modifications to **DL**. In contrast, we describe transformations that yield classical **DL** theories. Since the last two approaches are also described in Section 5, they are only briefly introduced here.

[Boutilier, 1992b] uses the correspondence between a conditional $\alpha_r \rightarrow \beta_r$ of System **Z** and defaults of the form $\frac{:\alpha_r \supset \beta_r}{\alpha_r \supset \beta_r}$ to produce partitioned sets of default rules. For rules in System **Z**, there is a corresponding set of prerequisite-free normal defaults. One can reason in **DL** by applying the rules in the highest set, and working down. Again, however, specificity is obtained by meta-theoretic considerations, in that one steps outside the machinery of **DL**. Also the order in which defaults are applied depends on the original **Z**-order; this order may be "upset" by the addition of irrelevant conditionals.

[Baader and Hollunder, 1993a] addresses specificity in terminological reasoners. In contrast to the present work, this approach does not rely on conflicts between "levels". Rather a subsumption relation between terminological concepts is mapped onto a set of partially ordered defaults in **DL**. [Brewka, 1993] has adopted the idea of minimal conflicting sets described here, but in a more restricted setting. In common with [Baader and Hollunder, 1993a], partially ordered defaults in **DL** are used; however, for inferencing all consistent strict total orders of defaults must be considered.

3 The Approach: Intuitions

As described previously, information in a **Z**-ordering is used to generate a default theory: The **Z**-ordering provides specificity information, and so for example, tells us that $P \rightarrow \neg F$ is a more specific rule than $B \rightarrow F$. However, we do not use the full **Z**-ordering, since it may introduce unwanted specificities (see Section 2.1). Rather we determine minimal sets of rules that conflict, and use these sets to sort out specificity information. The generated default theory (in **DL**) will be such that some inferences will be blocked (and so a penguin does not fly), while other inferences will go through (and so, penguins have wings).

Consider for example the following theory, already expressed as a **Z**-ordering:

$$\begin{aligned} R_0 &= \{An \rightarrow WB, An \rightarrow \neg Fe, An \rightarrow M\} \\ R_1 &= \{B \rightarrow An, B \rightarrow F, B \rightarrow Fe, B \rightarrow W\} \\ R_2 &= \{P \rightarrow B, P \rightarrow \neg F, E \rightarrow B, E \rightarrow \neg F, \\ &\qquad Pt \rightarrow B, Pt \rightarrow \neg Fe, Pt \rightarrow \neg WB\} \end{aligned}$$

That is, in R_0, animals are warm-blooded, don't have feathers, but are mobile. In R_1, birds are animals that fly, have feathers, and have wings. In R_2, penguins and emus are birds that don't fly, and pterodactyls are birds that have no feathers and are not warm-blooded.

First we locate the minimal sets of conditionals, such that there is a non-trivial **Z**-ordering for this set of conditionals. In our example these consist of:

$$\begin{aligned} C^0 &= \{An \rightarrow \neg Fe, B \rightarrow An, B \rightarrow Fe\} \\ C^1 &= \{B \rightarrow F, P \rightarrow B, P \rightarrow \neg F\} \qquad (6) \\ C^2 &= \{B \rightarrow F, E \rightarrow B, E \rightarrow \neg F\} \qquad (7) \\ C^3 &= \{B \rightarrow Fe, Pt \rightarrow B, Pt \rightarrow \neg Fe\} \\ C^4 &= \{An \rightarrow WB, B \rightarrow An, Pt \rightarrow B, Pt \rightarrow \neg WB\} \end{aligned}$$

Any such set is called a *minimal conflicting set* (*MCS*) of defaults. Such a set has a non-trivial **Z**-ordering, but for any subset there is no non-trivial **Z**-ordering. What this in turn means is that if all the rules in such a set are jointly applicable, then, one way or another there will be a conflict.[5] We show below that each such **Z**-ordering of a set C consists of a binary partition (C_0, C_1); furthermore the rules in the set C_0 are less specific than those in C_1. Consequently, if the rules in C_1 are applicable, then we would want to insure that some rule in C_0 was blocked.

Hence, for our initial example (1), we obtain one *MCS*, corresponding to (6), with the following **Z**-order:

$$\begin{aligned} C_0 &= \{B \rightarrow F\} \qquad (8) \\ C_1 &= \{P \rightarrow B, P \rightarrow \neg F\} \qquad (9) \end{aligned}$$

So there are two issues that need to be addressed:

1. What rules should be selected as candidates to be blocked?
2. How can the application of a rule be blocked?

For the first question, it turns out that there are different ways in which we can select rules. However, arguably the selection criterion should be independent of the default theory in which the rules are embedded, in the following fashion. For default theories R and R', where $R \subseteq R'$, if $r \in R$ is selected, then r should also be selected in R'. Thus, if we wish to block the default $B \rightarrow F$ in the case of P in default theory R, then we will also want to block this rule in any superset R'. In the sequel, we do this as follows: For a *MCS* C, we select those defaults in C_0 and C_1 that actually conflict and hence cause the non-triviality of C. The rules selected in this way from C_0 and C_1 are referred to as the *minimal conflicting rules* and *maximal conflicting rules* respectively. Then, the minimal conflicting rules constitute the candidates to be blocked.

Consider where we have a chain of rules, and where transitivity is explicitly blocked, such as may be found in an inheritance network:

$$A \rightarrow B_1, B_1 \rightarrow B_2, \ldots, B_n \rightarrow C \quad \text{but} \quad A \rightarrow \neg C.$$

[5] If the rules were represented as normal default rules in **DL** for example, one would obtain multiple extensions.

In this case, given A we need only block some rule in $A \rightarrow B_1, B_1 \rightarrow B_2, \ldots, B_n \rightarrow C$ to ensure that we do not obtain an inference from A to C. However, things are typically not so simplistic. Consider instead the *MCS* C^4 from above, expressed as a **Z**-ordering:

$$\begin{aligned} C_0 &= \{An \rightarrow WB, B \rightarrow An\} \\ C_1 &= \{Pt \rightarrow B, Pt \rightarrow \neg WB\} \end{aligned} \tag{10}$$

Intuitively An is less specific than Pt. Hence if we were given that An, Pt, $\neg B$ were true, then in a translation into default logic, we would want the default rule corresponding to $Pt \rightarrow \neg WB$ to be applicable over $An \rightarrow WB$, even though the "linking" rule $Pt \rightarrow B$ has been falsified. This in turn means that, for a *MCS*, we want the more specific rules to be applicable over the less specific conflicting rules, independently of the other rules in the *MCS*. We do this by locating those rules whose joint applicability would lead to an inconsistency. In the above, this would consist of $An \rightarrow WB$, and $Pt \rightarrow \neg WB$ since $An \wedge WB \wedge Pt \wedge \neg WB$ is inconsistent. Also we have that $An \rightarrow WB \in C_0$ and $Pt \rightarrow \neg WB \in C_1$ and so the rules have differing specificity.

For the second question, we have the following translation of rules into **DL**: The default theory corresponding to R consists of normal defaults, except for those defaults representing minimal conflicting rules, which will be semi-normal. For these latter default rules, the prerequisite is the antecedent of the original rule (as expected). The justification consists of the consequent together with an assertion to the effect that the maximal conflicting rules in the *MCS* hold.

Consider the set C_0 in (10), along with its minimal conflicting rule $An \rightarrow WB$. We replace $B \rightarrow An$, $Pt \rightarrow B$, $Pt \rightarrow \neg WB$ with $\frac{B : An}{An}$, $\frac{Pt : B}{B}$, $\frac{Pt : \neg WB}{\neg WB}$ respectively. For $An \rightarrow WB$, we replace it with

$$\frac{An : WB \wedge (Pt \supset \neg WB)}{WB},$$

which can be simplified to $\frac{An : WB \wedge \neg Pt}{WB}$. The rule $An \rightarrow WB$ is translated into a semi-normal default since it is the (only) minimal conflicting rule of C^4 (and of no other C^i). On the other hand, the rule $Pt \rightarrow WB$ is translated into a normal default since it does not occur as a minimal conflicting rule elsewhere.

So, for the minimal conflicting rules we obtain semi-normal defaults; all other defaults are normal. Accordingly, we give below only the semi-normal default rules constructed from the *MCS*s C^0, C^1, C^2, and C^3:

$$\begin{array}{lccc} C^0 : & \frac{An : \neg Fe \wedge \neg B}{\neg Fe} & & \\ C^1 + C^2 : & \frac{B : F \wedge (P \supset \neg F) \wedge (E \supset \neg F)}{F} & \text{or} & \frac{B : F \wedge \neg P \wedge \neg E}{F} \\ C^3 : & \frac{B : Fe \wedge (Pt \supset \neg Fe)}{Fe} & \text{or} & \frac{B : Fe \wedge \neg Pt}{Fe}. \end{array}$$

The conditional $B \rightarrow F$ occurs in C^1 and C^2 as a minimal conflicting rule. In this case we have two *MCS*s sharing the same minimal conflicting rule, and we combine the maximal conflicting rules of both sets.

So why does this approach work? The formal details are given in the following sections. However, informally, consider first where we have a *MCS* of defaults C with a single minimal conflicting rule $\alpha_0 \rightarrow \beta_0$ and a single maximal conflicting rule $\alpha_1 \rightarrow \beta_1$. If we are able to prove that α_0 (and so in **DL** can prove the antecedent of the conditional), then we would want β_0 to be a default conclusion—provided that no more specific rule applies. But what should constitute the justification? Clearly, first that β_0 is consistent. But also that "appropriate", more specific, conflicting conditionals not be applicable. Hence we add these more specific conditionals as part of the justification. Now, in our simplified setting, $\alpha_0 \rightarrow \beta_0$ is such that $\{\alpha_0 \wedge \beta_0\}$ is satisfiable, but for the conditional $\alpha_1 \rightarrow \beta_1$, $\{\alpha_0 \wedge \beta_0\} \cup \{\alpha_1 \wedge \beta_1\}$ is unsatisfiable. Hence it must be that $\{\alpha_0 \wedge \beta_0\} \cup \{\alpha_1 \supset \beta_1\} \models \neg \alpha_1$ for these conditionals. Thus if a minimal conflicting rule is applicable, then the maximal rule cannot be applicable.

This suggests that we might simply add the negation of the antecedent of the higher-level conflicting conditional. However the next example illustrates that this strategy does not work whenever a *MCS* has more than one minimal conflicting rule. Consider for example the following theory, already expressed as a **Z**-ordering:

$$R_0 = \{A \rightarrow \neg B, C \rightarrow \neg D\} \tag{11}$$

$$R_1 = \{A \wedge C \rightarrow B \vee D\} \tag{12}$$

If we were to represent this as a normal default theory, then with $\{A, C\}$ we would obtain three extensions, containing $\{\neg B, D\}$, $\{B, \neg D\}$, $\{\neg B, \neg D\}$. The last extension is unintuitive since it prefers the two less specific rules over the more specific one in R_1.

Now observe that the rules in $R_0 \cup R_1$ form a *MCS* with two minimal conflicting rules. In our approach, this yields two semi-normal defaults[6]

$$\begin{array}{cccc} \frac{A : \neg B \wedge (A \wedge C \supset B \vee D)}{\neg B} & \text{or} & \frac{A : \neg B \wedge (C \supset D)}{\neg B} & \text{and} \\ \frac{C : \neg D \wedge (A \wedge C \supset B \vee D)}{\neg D} & \text{or} & \frac{C : \neg D \wedge (A \supset B)}{\neg D} & \end{array}$$

along with the normal default rule $\frac{A \wedge C : B \vee D}{B \vee D}$. Given $\{A, C\}$, we obtain only the two more specific extensions, containing $\{\neg B, D\}$ and $\{B, \neg D\}$. In both cases, we apply the most specific rule, along with one of the less specific rules.

Note that if we add either only the negated antecedent of the maximal conflicting rule (viz. $\neg A \vee \neg C$) or all remaining rules (e.g. $C \supset \neg D$ and $A \wedge C \supset B \vee D$ in the case of the first default) to the justification of the two semi-normal defaults, then in both cases we obtain justifications that are too strong. For instance, for $A \rightarrow \neg B$ we would obtain either $\frac{A : \neg B \wedge (\neg A \vee \neg C)}{\neg B}$ which simplifies to $\frac{A : \neg B \wedge \neg C}{\neg B}$ or $\frac{A : \neg B \wedge (A \wedge C \supset B \vee D) \wedge (C \supset \neg D)}{\neg B}$

[6] We simplify justifications by replacing each occurrence of the prerequisite by true. The correctness for arbitrary prerequisites is shown in [Delgrande and Schaub, 1993].

which also simplifies to $\frac{A:\neg B\wedge\neg C}{\neg B}$. Given $\{A, C, D\}$ there is, however, no reason why the rule $A \rightarrow \neg B$ should not apply. In contrast, our construction yields the default $\frac{A:\neg B\wedge(C\supset D)}{\neg B}$, which blocks the second semi-normal default rule in a more subtle way, and additionally allows us to conclude $\neg B$ from $\{A, C, D\}$.

One can also show that conflicts that do not result from specificity (as found for example, in the "Nixon diamond") are handled correctly. These and other examples are discussed further following the presentation of the formal details.

4 Minimal Conflicting Sets

In what follows, we consider a **Z**-consistent set of default conditionals $R = \{r \mid \alpha_r \rightarrow \beta_r\}$ where each α_r and β_r are propositional formulas over a finite alphabet. We write *Prereq*(R) for $\{\alpha_r \mid \alpha_r \rightarrow \beta_r \in R\}$, and *Conseq*$(R)$ for $\{\beta_r \mid \alpha_r \rightarrow \beta_r \in R\}$.

For a set of rules R, the set of its *MCS*s represents conflicts among rules in R due to disparate specificity. Each *MCS* is a minimal set of conditionals having a non-trivial **Z**-ordering.

Definition 2 *Let R be a* **Z***-consistent set of rules. $C \subseteq R$ is a minimal conflicting set (MCS) in R iff C has a non-trivial* **Z***-ordering and any $C' \subset C$ has a trivial* **Z***-ordering.*

Observe that adding new rules to R cannot alter or destroy any existing *MCS*s. That is, for default theories R and R', where $C \subseteq R \subseteq R'$, we have that if C is a *MCS* in R then C is a *MCS* in R'.

The next theorem shows that any *MCS* has a binary partition:[7]

Theorem 1 *Let C be a MCS in R. Then, the* **Z***-ordering of C is (C_0, C_1) for some non-empty sets C_0 and C_1 with $C = C_0 \cup C_1$.*

Moreover, a *MCS* entails the negations of the antecedents of the higher-level rules:

Theorem 2 *Let C be a MCS in R. Then, if $\alpha \rightarrow \beta \in C_1$ then $C \models \neg\alpha$.*

Hence, given the rule set in (1),

$$R = \{B \rightarrow F,\ B \rightarrow W,\ P \rightarrow B,\ P \rightarrow \neg F\},$$

there is one *MCS*

$$C = \{B \rightarrow F,\ P \rightarrow B,\ P \rightarrow \neg F\}.$$

As shown in (8/9), the first conditional constitutes C_0 and the last two C_1 in the **Z**-order of C. The set $\{B \rightarrow F,\ P \rightarrow \neg F\}$ for example, is not a *MCS* since alone it has a trivial **Z**-order. It is easy to see that $C \models \neg P$.

[7] Proofs are omitted for space limitations, but can be found in [Delgrande and Schaub, 1993].

Intuitively, a *MCS* consists of three mutually exclusive sets of rules: the least specific or *minimal conflicting rules* in C, *min*(C); the most specific or *maximal conflicting rules* in C, *max*(C); and the remaining rules providing a minimal inferential relation between these two sets of rules, *inf*(C). The following definition provides a very general formal frame for these sets:

Definition 3 *Let R be a set of rules and let $C \subseteq R$ be a MCS in R. We define max(C) and min(C) to be non-empty subsets of R such that*

$$\begin{aligned} \mathit{min}(C) &\subseteq C_0 \\ \mathit{max}(C) &\subseteq C_1 \\ \mathit{inf}(C) &= C - (\mathit{min}(C) \cup \mathit{max}(C)) \end{aligned}$$

We observe that *min*, *max*, and *inf* are exclusive subsets of C such that $C = \mathit{min}(C) \cup \mathit{inf}(C) \cup \mathit{max}(C)$. We show below that the rules in *max*(C) and *min*(C) are indeed conflicting due to their different specificity. Note however that the following three theorems are independent of the choice of *min*(C), *inf*(C), and *max*(C). Yet after these theorems we argue in Definition 4 for a specific choice for these sets that complies with the intuitions described in the previous section.

First, the antecedents of the most specific rules in *min*(C) imply the antecedents of the least specific rules in *max*(C) modulo the "inferential rules":

Theorem 3 *Let C be a MCS in a set of rules R. Then, $\mathit{inf}(C) \cup \mathit{max}(C) \models \mathit{Prereq}(\mathit{max}(C)) \supset \mathit{Prereq}(\mathit{min}(C))$.*

In fact, $\mathit{inf}(C) \cup \mathit{max}(C)$ is the weakest precondition under which the last entailment holds. This is important since we deal with a general setting for *MCS*s. Observe that omitting *max*(C) would eliminate rules that may belong to *max*(C), yet provide "inferential relations". The next theorem shows that the converse of the previous does not hold in general.

Theorem 4 *Let C be a MCS in a set of rules R. Then, for any set of rules R' such that $C \subseteq R'$ and any set of rules $R'' \subseteq \mathit{min}(C)$ such that $R' \cup \mathit{Prereq}(R'')$ is satisfiable, we have: $R' \not\models \mathit{Prereq}(R'') \supset \mathit{Prereq}(\mathit{max}(C))$.*

The reason for considering consistent subsets of *min*(C) is that its entire set of prerequisites might be equivalent to those in *max*(C). Then, however, $C \cup \mathit{Prereq}(\mathit{min}(C))$ and so $R' \cup \mathit{Prereq}(\mathit{min}(C))$ is inconsistent. This is, for instance, the case in Equation (11/12). In fact, R' is the strongest precondition under which the above theorem holds. Finally, we demonstrate that these rules are indeed conflicting.

Theorem 5 *Let C be a MCS in a set of rules R. Then, for any $\alpha \rightarrow \beta \in \mathit{max}(C)$, we have: $\mathit{inf}(C) \cup \{\alpha\} \models \neg(\mathit{Conseq}(\mathit{min}(C)) \wedge \mathit{Conseq}(\mathit{max}(C)))$.*

As above, $\mathit{inf}(C) \cup \{\alpha\}$ is the weakest precondition under which the last entailment holds. In all, the last three theorems demonstrate that the general framework given for *MCS*s (already) provides an extremely expressive way of isolating rule conflicts due to their specificity.

4.1 Specific Minimal and Maximal Conflicting Rules

As indicated in Section 3, we require further restrictions on the choice of $min(C)$ and $max(C)$ for our translation into **DL**. For a *MCS* $C = (C_0, C_1)$, we have the information that the rules in C_0 are less specific than those in C_1. However we wish to isolate those rules in C_0 whose application would conflict with applications of rules in C_1. Such a set is referred to as a *conflicting core* of a *MCS*. This leads us to the following definition:

Definition 4 *Let $C = (C_0, C_1)$ be a MCS. A conflicting core of C is a pair of least non-empty sets $(min(C), max(C))$ where*

1. $min(C) \subseteq C_0$,
2. $max(C) \subseteq C_1$,
3. $\{\alpha_r \wedge \beta_r \mid r \in max(C) \cup min(C)\} \models \perp$.

This definition specializes the general setting of Definition 3. So, $\alpha_r \rightarrow \beta_r$ is in $min(C)$ if its application conflicts with the application of a rule (or rules) in C_1.

In the extended example of Section 3 the conflicting cores are

$$\begin{array}{ll} C^0: & (\{An \rightarrow \neg Fe\}, \{B \rightarrow Fe\}) \\ C^1: & (\{B \rightarrow F\}, \{P \rightarrow \neg F\}) \\ C^2: & (\{B \rightarrow F\}, \{E \rightarrow \neg F\}) \\ C^3: & (\{B \rightarrow Fe\}, \{Pt \rightarrow \neg Fe\}) \\ C^4: & (\{An \rightarrow WB\}, \{Pt \rightarrow \neg WB\}) \end{array}$$

respectively. The conflicting core of our initial example in (1) corresponds to the one for C^1. For a complement consider the example given in (11/12), where the conflicting core contains two minimal and one maximal conflicting rules:

$$(\{A \rightarrow \neg B, C \rightarrow \neg D\}, \{A \wedge C \rightarrow B \vee D\}).$$

Note that a conflicting core need not necessarily exist for a specific *MCS*. For example, consider the *MCS* (expressed as a **Z**-order):

$$\begin{array}{rcl} C_0 & = & \{Q \rightarrow P, R \rightarrow \neg P\} \\ C_1 & = & \{Q \wedge R \rightarrow PA\} \end{array}$$

Thus Quakers are pacifists while republicans are not; Quakers that are republicans are politically active. Here the conflict is between two defaults at the same level (viz. $Q \rightarrow P$ and $R \rightarrow \neg P$) that manifests itself when a more specific default is given.

We do have the following result however.

Theorem 6 *For MCS C in a set of rules R, if $\{\alpha_r \wedge \beta_r \mid r \in min(C)\} \not\models \perp$ and $\{\alpha_r \wedge \beta_r \mid r \in max(C)\} \not\models \perp$ then C has a conflicting core.*

5 Compiling Specificity into Default Theories

In the previous section, we proposed an approach for isolating minimal sets of rules that conflict because of their different specificity. We also showed how to isolate specific minimal and maximal rules. In this section, we use this information for specifying blocking conditions or, more generally, priorities among conflicting defaults in Default Logic. To this end, we envisage two different possible approaches. First, we could determine a strict partial order on a set of rules R from the *MCS*s in R. That is, for two rules $r, r' \in R$, we can define $r < r'$ iff $r \in min(C)$ and $r' \in max(C)$ for some *MCS* C in R. In this way, $r < r'$ is interpreted as "r is less specific than r'". Then, one could interpret each rule $\alpha \rightarrow \beta$ in R as a normal default $\frac{\alpha : \beta}{\beta}$ and use one of the approaches developed in [Baader and Hollunder, 1993a] or [Brewka, 1993] for computing the extensions of ordered normal default theories, i.e. default theories enriched by a strict partial order on rules. These approaches however have the disadvantage that they step outside the machinery of **DL** for computing extensions.

This motivates an alternative approach that remains inside the framework of classical **DL**, where we automatically transform rules with specificity information into semi-normal default theories.

5.1 Z-Default Logic

This section describes a strategy, based on the notions of specificity and conflict developed in the previous section, for producing a standard semi-normal default theory, and which provable maintains this notion of specificity. The transformation is succinctly defined:

Definition 5 *Let R be a set of rules and let $\langle C^i \rangle_{i \in I}$ be the family of all MCSs in R. For each $r \in R$, we define*

$$\delta_r = \frac{\alpha_r \, : \, \beta_r \wedge \bigwedge_{r' \in R_r} (\alpha_{r'} \supset \beta_{r'})}{\beta_r} \qquad (13)$$

where $R_r = \{r' \in max(C^i) \mid r \in min(C^i) \text{ for } i \in I\}$. We define $D_R = \{\delta_r \mid r \in R\}$.

In what follows, we adopt the latter notation and write $D_{R'} = \{\delta_r \mid r \in R'\}$ for any subset R' of R.

The most interesting point in the preceding definition is the formation of the justifications of the (sometimes) semi-normal defaults. Given a rule r, the justification of δ_r is built by looking at all *MCS*, C^i, in which r occurs as a least specific rule (i.e. $r \in min(C^i)$). Then, the consequent of r is conjoined with the strict counterparts of the most specific rules in the same sets (viz. $(\alpha_{r'} \supset \beta_{r'})$ for $r' \in max(C^i)$). Hence, for the minimal conflicting rules we obtain semi-normal defaults; all other defaults are normal (since then $R_r = \emptyset$). So for any *MCS* C in R, we transform the rules in $min(C)$ into semi-normal defaults, whereas we transform the rules in $inf(C) \cup max(C)$ into normal defaults, provided that they do not occur elsewhere as a minimal conflicting rule.

As suggested in Section 4.1, we are only interested in minimal and maximal conflicting rules forming a conflicting core. That is, given a *MCS* C, we stipulate that $(min(C), max(C))$ forms a conflicting core of C. In the extended example of Section 3 the conflicting cores for (6) and (7) are

$$(\{B \to F\}, \{P \to \neg F\}) \text{ and } (\{B \to F\}, \{E \to \neg F\})$$

respectively. According to Definition 5, we get $R_{B\to F} = \{P \to \neg F, E \to \neg F\}$. This results in a single semi-normal default rule

$$\frac{B : F \wedge (P \supset \neg F) \wedge (E \supset \neg F)}{F}, \qquad \text{or} \qquad \frac{B : F \wedge \neg P \wedge \neg E}{F}.$$

Observe that we obtain $\frac{P:B}{B}$ and $\frac{P:\neg F}{\neg F}$ for $P \to B$, and $P \to \neg F$ since these rules do not occur elsewhere as minimal rules in a conflicting core. Other examples were given at the end of Section 3.

For a more general example, consider the case where, given a rule r, R_r is a singleton set containing a rule r'. Thus r is less specific than r'. This results in the default rules $\frac{\alpha_r : \beta_r \wedge (\alpha_{r'} \supset \beta_{r'})}{\beta_r}$ and $\frac{\alpha_{r'} : \beta_{r'}}{\beta_{r'}}$. Our intended interpretation is that r and r' conflict, and that r is preferable over r' (because of specificity). Thus, assume that β_r and $\beta_{r'}$ are not jointly satisfiable. Then, the second default takes precedence over the first one, whenever both prerequisites are derivable (i.e. $\alpha_r \in E$ and $\alpha_{r'} \in E$), and both β_r and $\beta_{r'}$ are individually consistent with the final extension E (i.e. $\neg\beta_r \notin E$ and $\neg\beta_{r'} \notin E$). That is, while the justification of the second default is satisfiable, the justification of the first default, $\beta_r \wedge (\alpha_{r'} \supset \beta_{r'})$, is unsatisfiable.

In general, we obtain the following results. $GD(E, D)$ stands for the generating defaults of E with respect to D, i.e. $GD(E, D) = \{\frac{\alpha : \beta}{\omega} \in D, \mid \alpha \in E, \neg\beta \notin E\}$. Note that Theorem 7 is with respect to the general theory of *MCS*s while Theorem 8 is with respect to the specific development involving conflicting cores.

Theorem 7 *Let R be a set of rules and let W be a set of formulas. Let C be a MCS in R. Let E be a consistent extension of (D_R, W). Then,*

1. *if $D_{max(C)} \cup D_{inf(C)} \subseteq GD(E, D)$ then $D_{min(C)} \not\subseteq GD(E, D)$,*
2. *if $D_{min(C)} \cup D_{inf(C)} \subseteq GD(E, D)$ then $D_{max(C)} \not\subseteq GD(E, D)$.*

Let us relate this theorem to the underlying idea of specificity: Observe that in the first case, where $D_{max(C)} \cup D_{inf(C)} \subseteq GD(E, D)$, we also have

$$Prereq(min(C)) \subseteq E$$

by Theorem 3. That is, even though the prerequisites of the minimal conflicting defaults are derivable, they do not contribute to the extension at hand. This is so because some of the justifications of the minimal conflicting defaults are not satisfied. In this way, the more specific defaults in $D_{max(C)}$ take precedence over the less specific defaults in $D_{min(C)}$. Conversely, in the second case, where $D_{min(C)} \cup D_{inf(C)} \subseteq GD(E, D)$, the less specific defaults apply only if the more specific defaults do not contribute to the given extension.

Theorem 8 *Let R be a set of rules and let W be a set of formulas. Let $(min(C), max(C))$ be a conflicting core of some MCS C in R. Let E be a consistent extension of (D_R, W). Then,*

1. *if $D_{max(C)} \subseteq GD(E, D)$ then $D_{min(C)} \not\subseteq GD(E, D)$,*
2. *if $D_{min(C)} \subseteq GD(E, D)$ then $D_{max(C)} \not\subseteq GD(E, D)$.*

Thus in this case we obtain that the defaults in a conflicting core are not applicable, independent of the "linking defaults" in $D_{inf(C)}$.

Given a set of formulas W representing our world knowledge and a set of default conditionals R, we can apply Definition 5 in order to obtain a so-called **Z**-*default theory* (D_R, W). The following theorem gives an alternative characterization for extensions of **Z**-default theories. In particular, it clarifies further the effect of the set of rules R_r associated with each rule r. Recall that in general, however, such extensions are computed in the classical framework of **DL**.

Theorem 9 *Let R be a set of rules, let $D_N = \left\{ \frac{\alpha_r : \beta_r}{\beta_r} \;\middle|\; \alpha_r \to \beta_r \in R \right\}$, and let W and E be sets of formulas. Define $E_0 = W$ and for $i \geq 0$ (and R_r as in Definition 5)*

$$E_{i+1} = Th(E_i) \cup \left\{ \beta_r \;\middle|\; \frac{\alpha_r : \beta_r}{\beta_r} \in D_N,\ \alpha_r \in E_i, \right. \left. E \cup \{\beta_r\} \cup \textstyle\bigcup_{r' \in R_r} (\alpha_{r'} \supset \beta_{r'}) \not\vdash \bot \right\}$$

Then, E is an extension of (D_R, W) iff $E = \bigcup_{i=0}^{\infty} E_i$.

5.2 Properties of Z-Default Theories

We now examine the formal properties of **Z**-default theories. In regular **DL**, many appealing properties are only enjoyed by restricted subclasses. For instance, normal default theories guarantee the existence of extensions and enjoy the property of semi-monotonicity.

Transposed to our case, the latter stipulates that if $R' \subseteq R$ for two sets of rules, then if E' is an extension of $(D_{R'}, W)$ then there is an extension E of

(D_R, W) where $E' \subseteq E$. Arguably, this property is not desirable if we want to block less specific defaults in the presence of more specific defaults. In fact, this property does not hold for **Z**-default theories. For instance, from the rules $B \rightarrow F, P \rightarrow B$, we obtain the defaults $\frac{B:F}{F}, \frac{P:B}{B}$. Given P, we conclude B and F. However, adding the rule $P \rightarrow \neg F$ makes us add the default $\frac{P:\neg F}{\neg F}$ and replace the default $\frac{B:F}{F}$ by $\frac{B:F\wedge\neg P}{F}$. Obviously, the resulting theory does not support our initial conclusions. Rather we conclude now B and $\neg F$, which violates the aforementioned notion of semi-monotonicity.[8]

Also, the existence of extensions is not guaranteed for **Z**-default theories. To see this, consider the rules:

$$\begin{array}{lllllllll} A\wedge Q & \rightarrow & \neg P & B\wedge R & \rightarrow & \neg Q & C\wedge P & \rightarrow & \neg R \\ A & \rightarrow & P & B & \rightarrow & Q & C & \rightarrow & R \end{array}$$

Each column gives a *MCS* in which the upper rule is more specific than the lower rule. We obtain the rules

$$\begin{array}{ccc} \frac{A\wedge Q:\neg P}{\neg P} & \frac{B\wedge R:\neg Q}{\neg Q} & \frac{C\wedge P:\neg R}{\neg R} \\ \frac{A:P\wedge\neg Q}{P} & \frac{B:Q\wedge\neg R}{Q} & \frac{C:R\wedge\neg P}{R} \end{array}$$

Given A, B, C, we get no extension.

Arguably, the non-existence of extensions indicates certain problems in the underlying set of rules. [Zhang and Marek, 1990] shows that a default theory has no extension iff it contains certain "abnormal" defaults; these can be detected automatically. However, we can also avoid the non-existence of extensions by translating rules into variants of default logic that guarantee the existence of extensions, as discussed in Section 6.

Another important property is cumulativity. The intuitive idea is that if a theorem is added to the set of premises from which the theorem was derived, then the set of derivable formulas should remain unchanged. This property is only enjoyed by prerequisite-free normal default theories in regular **DL**. It does not hold for **Z**-default theories, as the next example illustrates. Consider the conditionals $\{D \rightarrow A, A \rightarrow B, B \rightarrow \neg A\}$. The last two conditionals form a *MCS*. Transforming these rules into defaults, yields two normal, $\frac{D:A}{A}, \frac{A:B}{B}$, and one (semi-)normal default, $\frac{B:\neg A\wedge(A\supset B)}{\neg A}$, or $\frac{B:\neg A}{\neg A}$. Given D, there is one extension containing $\{D, A, B\}$. Hence this extension contains B. Now, given D and B, we obtain a second extension containing $\{D, \neg A, B\}$. This violates cumulativity.

Note that in this case we obtained a normal default theory from the original set of rules. This is intuitively plausible, since the two conflicting defaults are mutually canceling, i.e. if one applies then the other does not.

[8]This differs from the notion of semi-monotonicity described in [Reiter, 1980]. The latter is obtained by replacing R and D_R by D and R' and $D_{R'}$ by D'.

5.3 Exchangeability and Related Work

At the start of this section we described how to extract a strict partial order from a family of *MCS*s for using other approaches (such as [Baader and Hollunder, 1993a; Brewka, 1993]) to compute extensions of ordered default theories, i.e. theories with a strict partial order $<$ on the defaults. In fact, one can view partial orders on rules as general interfaces between approaches. In particular, we can use also our approach for compiling ordered normal default theories into semi-normal default theories. To this end, we have to incorporate the order $<$ into the specification of R_r in Definition 5. We do this by associating with each normal default $\frac{\alpha:\beta}{\beta}$ a rule $\alpha \rightarrow \beta$ and define for each such rule r that $R_r^< = \{r' \mid r < r'\}$, where $<$ is a strict partial order on the set of rules. Then, we can use transformation (13) for turning ordered normal default theories into semi-normal default theories.

We can now compare how priorities are dealt with in our and the aforementioned approaches. In both [Baader and Hollunder, 1993a] and [Brewka, 1993] the iterative specification of an extension in **DL** is modified. In brief, a default is only applicable at an iteration step if no more specific (or $<$-greater) default is applicable.[9] The difference between both approaches (roughly) rests on the number of defaults applicable at each step. While Brewka allows only for applying a single default that is maximal with respect to a total extension of $<$, Baader and Hollunder allow for applying all $<$-maximal defaults at each step.

As a first example, consider the default rules $\frac{:A}{A}, \frac{:B}{B}, \frac{B:C}{C}, \frac{A:\neg C}{\neg C}$ (for short $\delta_1, \delta_2, \delta_3, \delta_4$), along with $\delta_4 < \delta_3$, taken from [Baader and Hollunder, 1993b]. With no facts Baader and Hollunder obtain one extension containing $\{A, B, C\}$. Curiously, Brewka obtains an additional extension containing $\{A, B, \neg C\}$. In our approach, we generate from $<$ a single nonempty set $R_{\delta_4}^< = \{\delta_3\}$; all other such sets are empty. Consequently we replace δ_4 by $\frac{A:\neg C\wedge(B\supset C)}{\neg C}$ or $\frac{A:\neg C\wedge\neg B}{\neg C}$. In regular **DL**, the resultant default theory yields only the first extension containing $\{A, B, C\}$.

As a second example, again from [Baader and Hollunder, 1993b], consider the rules $\frac{:A}{A}, \frac{B:\neg A}{\neg A}, \frac{:B}{B}, \frac{A:\neg B}{\neg B}$ (for short $\delta_1, \delta_2, \delta_3, \delta_4$), along with $\delta_1 < \delta_2$, $\delta_3 < \delta_4$. They show that in Brewka's approach two extensions are obtained, one containing $\{A, \neg B\}$ and another containing $\{\neg A, B\}$. However an additional extension is obtained in Baader and Hollunder's approach, containing $\{A, B\}$. In our approach, we produce from $<$ the nonempty sets $R_{\delta_1}^< = \{\delta_2\}$; and $R_{\delta_3}^< = \{\delta_4\}$; all other such sets are empty. Then, we replace δ_1 and δ_3 by

$$\frac{:A\wedge(B\supset\neg A)}{A} \text{ or } \frac{:A\wedge\neg B}{A} \quad \text{and} \quad \frac{:B\wedge(A\supset\neg B)}{B} \text{ or } \frac{:B\wedge\neg A}{B},$$

[9]In [Baader and Hollunder, 1993a; Brewka, 1993] $<$ is used in the reverse order.

which yields only the first two extensions in **DL**.

Even though these examples appear to be artificial, they can be extended to express reasonable specificity orderings. In all, we observe that in both examples our approach yields the fewer and, in terms of specificity, more intuitive extensions.

Note that the general approach of compiling partial orders into semi-normal default theories makes sense whenever we deal with partial orders that only consider priorities due to specificity where we have *truly conflicting rules*. Otherwise, the resulting default theory may be overly strong. Consider the case where we extract priorities from subsumption relations, as is done in [Baader and Hollunder, 1993a] for terminological logics. Consider terms stating that "birds fly", $B \rightarrow F$, and "young birds need special care", $Y \rightarrow C$, along with the usual subsumption relation between "birds" and "young birds". This subsumption amounts to a priority between the two rules even though there is no conflict: $(B \rightarrow F) < (Y \rightarrow C)$. Thus these rules would result in two default rules $\frac{B : F \wedge (Y \supset C)}{F}$ and $\frac{Y : C}{C}$ since the first default would "take priority" over the second, according to the given partial order. Such a priority is unnecessary however as regards avoiding conflicts stemming from more specific information. Obviously, this problem does not arise in the general approach taken by *MCS*s. In this case, in addition to a specificity difference, we also require explicitly conflicting rules. In the above example there is no *MCS* and so we would obtain the two normal rules $\frac{B : F}{F}$ and $\frac{Y : C}{C}$.

Finally we note that the preceding exposition was dominated by the view that rules, like $\alpha \rightarrow \beta$, are associated with defaults having prerequisite α and consequent β. This view underlies the approaches in [Baader and Hollunder, 1993a] and [Brewka, 1993]. That is, they rely on the existence of prerequisites. In contrast, we can treat rules also as strict implications, and so compile them into a prerequisite-free defaults, as we show in the next section.

6 Alternative Translations

So far we have focused on translating specificity information into Reiter's default logic. In this section, we show how the specificity information extracted from a family of minimal conflicting sets (or even a strict partial order) can be incorporated into alternative approaches to default reasoning.

As mentioned earlier, we can also interpret a rule $\alpha \rightarrow \beta$ as a strict implication, namely $\alpha \supset \beta$. To this end, we turn rules like $\alpha \rightarrow \beta$ into prerequisite-free default rules. However, as discussed in [Delgrande *et al.*, 1994], the problem of controlling interactions among such rules is more acute than in the regular case. Consider our initial example (1), translated into prerequisite-free **DL**:

$$\frac{: B \supset F}{B \supset F}, \quad \frac{: B \supset W}{B \supset W}, \quad \frac{: P \supset B}{P \supset B}, \quad \frac{: P \supset \neg F}{P \supset \neg F} \tag{14}$$

Given P, we obtain three extensions, containing $\{P, \neg F, B, W\}$, $\{P, F, B, W\}$, and $\{P, \neg F, \neg B\}$.[10] The first two extensions correspond to the ones obtained in regular **DL**. Clearly, we can apply the techniques developed in the previous sections for eliminating the second extension. The third extension yields also the more specific result in that we obtain $\neg F$. This extension, however, does not account for property inheritance, since we cannot conclude that birds have wings. This is caused by the contraposition of $B \supset F$. That is, once we have derived $\neg F$, we derive $\neg B$ by contraposition, which prevents us from concluding W.

This problem can be addressed in two ways, either by strenthening the blocking conditions for minimal conflicting rules or by blocking the contraposition of minimal conflicting rules. In the first case, we could turn $B \rightarrow F$ into $\frac{: (B \supset F) \wedge \neg P}{B \supset F}$ by adding the negated antecedants of the maximal conflicting rules, here $\neg P$. While this looks appealing, we have already seen in Section 3 that this approach is too strong in the presence of multiple minimal conflicting rules. To see this, consider the rules given in (11/12). For $A \rightarrow \neg B$, we would obtain $\frac{: (A \supset \neg B) \wedge (\neg A \vee \neg C)}{A \supset \neg B}$ or $\frac{: A \supset (\neg B \wedge \neg C)}{A \supset \neg B}$. However, as argued in Section 3, there is no reason why $A \rightarrow \neg B$ should not be applied given the facts $\{A, C, D\}$. Also, in general it does not make sense to address a problem stemming from contrapositions by altering the way specificity is enforced. Rather we should address an independent problem by means of other measures.

So, in the second case, we turn $B \rightarrow F$ into $\frac{: (B \supset F) \wedge F \wedge (P \supset \neg F)}{B \supset F}$ or $\frac{: F \wedge \neg P}{B \supset F}$. That is, we add the consequent of $B \rightarrow F$ in order to block its contraposition. As before, we add the strict counterparts of the maximal conflicting rules, here $P \supset \neg F$. In the birds example, the resulting justification is strengthened as above. In particular, we block the contribution of the rule $B \supset F$ to the final extension if either $\neg F$ or P is derivable. For $A \rightarrow \neg B$ in (11/12), we now obtain, $\frac{: (A \supset \neg B) \wedge \neg B \wedge (A \wedge C \supset B \vee D))}{A \supset \neg B}$ or $\frac{: \neg B \wedge (A \wedge C \supset D)}{A \supset \neg B}$. In contrast to the previous proposal, this rule is applicable to the facts $\{A, C, D\}$. Moreover, this approach is in accord with System **Z**, where rules are classified according to their "forward chaining" behaviour.

So for translating rules along with their specificity into prerequisite-free default theories, we replace the definition of δ_r in Definition 5 by[11]

$$\zeta_r = \frac{: (\alpha_r \supset \beta_r) \wedge \beta_r \wedge \bigwedge_{r' \in R_r} (\alpha_{r'} \supset \beta_{r'})}{(\alpha_r \supset \beta_r)}. \tag{15}$$

[10] The third extension would not be present if $P \supset B$ were a strict rule.

[11] Observe that $(\alpha_r \supset \beta_r) \wedge \beta_r$ is equivalent to β_r.

Applying this transformation to our birds example in (1), we obtain:

$$\frac{:F\wedge\neg P}{B\supset F},\ \frac{:B\supset W}{B\supset W},\ \frac{:P\supset B}{P\supset B},\ \frac{:P\supset\neg F}{P\supset\neg F}$$

Now, given P, we obtain a single extension containing $\{P, \neg F, B, W\}$.

Note that blocking the contraposition of minimal conflicting rules is an option outside the presented framework. The purpose of the above transformation is to preserve inheritance over default statements, like $P \rightarrow B$. Inheritance over strict statements, like $P \supset B$, however can be done *without* blocking contrapositions. In this case, the following transformation is sufficient:

$$\zeta'_r = \frac{:\ (\alpha_r \supset \beta_r) \wedge \bigwedge_{r'\in R_r}(\alpha_{r'} \supset \beta_{r'})}{(\alpha_r \supset \beta_r)} \qquad (16)$$

As an example, let us turn the default $P \rightarrow B$ into its strict counterpart $P \supset B$. As detailed in [Delgrande and Schaub, 1993], our birds example then yields with transformation (16) the defaults

$$\frac{:(B\supset F)\wedge(P\supset\neg F)}{B\supset F},\ \frac{:B\supset W}{B\supset W},\ \frac{:P\supset\neg F}{P\supset\neg F}.$$

Now, given P *and* $P \supset B$, we obtain a single extension containing $\{P, \neg F, B, W\}$. The details on integrating strict rules are given in [Delgrande and Schaub, 1993].

Transformations (15/16) offer some interesting benefits, since prerequisite-free defaults allow for reasoning by cases and reasoning by contraposition (apart from minimal conflicting rules). That is, such defaults behave like usual conditionals unless explicitly blocked. Nonetheless, the counterexamples for semi-monotonicity, cumulativity, and the existence of extensions carry over to prerequisite-free **Z**-default theories. Thus none of these properties is enjoyed by these theories in **DL**. Finally, note that this approach differs from [Boutilier, 1992b], where a ranking on defaults is obtained from the original **Z**-order; this may introduce unwanted priorities due to irrelevant conditionals.

Another alternative is the translation into variants of **DL** that guarantee the existence of extensions [Łukaszewicz, 1988; Brewka, 1991; Delgrande *et al.*, 1994]. This can be accomplished by means of both translation (13) and (15/16). Moreover, the resulting **Z**-default theories enjoy cumulativity when applying translation (13) and (15/16) in the case of Cumulative Default Logic and when applying translation (15/16) in the case of Constrained Default Logic. The corresponding results can be found in [Brewka, 1991; Delgrande *et al.*, 1994]. Although none of these variants enjoys semi-monotonicity with respect to the underlying conditionals, all of them enjoy this property with respect to the default rules. As shown in [Brewka, 1991], this may lead to problems in blocking a rule, like $\frac{B\,:\,F\wedge\neg P}{F}$, in the case $\neg P$ is a default conclusion. For details on this we refer the reader to [Brewka, 1991].

Similarly we can compile prioritized rules into Theorist [Poole, 1988] or other approaches, such as Autoepistemic Logic [Moore, 1985] or even Circumscription [McCarthy, 1980]. The latter translation is described in a forthcoming paper.

For the translation into Theorist, we refer the reader to [Delgrande *et al.*, 1994], where it is shown that Theorist systems correspond to prerequisite-free default theories in Constrained Default Logic and vice versa. Accordingly, we may obtain a Theorist system from a set of prioritized rules by first applying transformation (15/16) and then the one given in [Delgrande *et al.*, 1994] for translating prerequisite-free default theories in Constrained Default Logic into Theorist.

Autoepistemic Logic [Moore, 1985] aims at formalizing an agent's reasoning about her own beliefs. To this end, the logical language is augmented by a modal operator L. Then, a formula $L\alpha$ is to be read as "α is believed". For a set W of such formulas, an autoepistemic extension E is defined as

$$Th(W \cup \{L\alpha \mid \alpha \in E\} \cup \{\neg L\alpha \mid \alpha \notin E\}).$$

As discussed in [Konolige, 1988], we can express a statement like "birds fly" either as $B \wedge \neg L\neg F \supset F$ or $LB \wedge \neg L\neg F \supset F$. Given B and one of these rules, we obtain in both cases an extension containing F. Roughly speaking, the former sentence corresponds to the default $\frac{:B\supset F}{B\supset F}$ while the latter is close to $\frac{B\,:\,F}{F}$.

This motivates the following translations into Autoepistemic Logic. Let R be a set of rules and let $R_r \subseteq R$, for each $r \in R$ we define:

$$\rho_r = \alpha_r \wedge \neg L\neg\left(\beta_r \wedge \bigwedge_{r'\in R_r}(\alpha_{r'} \supset \beta_{r'})\right) \supset \beta_r,$$

$$\varrho_r = L\alpha_r \wedge \neg L\neg\left(\beta_r \wedge \bigwedge_{r'\in R_r}(\alpha_{r'} \supset \beta_{r'})\right) \supset \beta_r.$$

Applying the first transformation to our initial example, we obtain for $B \rightarrow F$ the modal sentence

$$B\wedge\neg L\neg(F\wedge(P\supset\neg F))\supset F \text{ or } B\wedge\neg L\neg(F\wedge\neg P)\supset F,$$

along with $B \wedge \neg L\neg W \supset W, P \wedge \neg L\neg B \supset B$, and $P \wedge \neg LF \supset \neg F$ for $B \rightarrow W, P \rightarrow B$, and $P \rightarrow \neg F$. Now, given P along with the four modal defaults, we obtain a single autoepistemic extension containing $\neg F$ and W. In this way, we have added specificity to Autoepistemic Logic while preserving inheritance.

7 Discussion

This paper has described a hybrid approach addressing the notion of specificity in default reasoning. We begin with a set of rules that express default conditionals, where the goal is to produce a default theory expressed in a "target" formalism, and where conflicts arising from differing specificities are resolved. The approach is to use the techniques of a weak system, as exemplified by System **Z**, to isolate minimal sets of conflicting defaults. From the specificity information intrinsic in these sets, a default theory in a target language (here primarily Default Logic) is derived. In our approach, the problems of weak systems, such as lack of adequate property inheritance and undesirable specificity relations, are avoided. In addition, difficulties inherent in stronger systems, in particular, lack of specificity, are addressed. In contrast to previous work, the approach avoids stepping outside the machinery of **DL**. Thus we do not obtain an explicit global partial order on default rules, but rather a classical default theory where local conflicts are resolved by semi-normal defaults.

This approach is modular, in that we separate the *determination* of conflicts from the *resolution* of conflicts among rules. Thus either module could be replaced by some other approach. For example, one could use an inheritance network to determine conflict relations and then use the mapping described in this paper to obtain a default theory. Alternately, conflicts could be determined using *MCS*s via System **Z**, and then an ordered default theory as described in [Baader and Hollunder, 1993a] could be generated. The approach may be seen as generalising that of [Reiter and Criscuolo, 1981]. Also, for example, [Etherington and Reiter, 1983] and [Brewka, 1993] may be seen as falling into the same general framework.

Acknowledgements

We would like to thank Gerd Brewka for many discussions on the topic. The first author was a visitor at York University and the University of Toronto while this work was being carried out. The first author also acknowledges support from the Natural Science and Engineering Research Council of Canada grant A0884, as well as the Institute for Robotics and Intelligent Systems (IRIS) in the Canadian Networks of Centres of Excellence Program. The second author was supported by the Commission of the European Communities under grant no. ERB4001GT922433.

References

Baader, F. and Hollunder, B. 1993a. How to prefer more specific defaults in terminological default logic. In *Proc. IJCAI-93*, Chambéry, Fr. 669–674.

Baader, F. and Hollunder, B. 1993b. How to prefer more specific defaults in terminological default logic. Technical Report RR-92-58, DFKI.

Boutilier, Craig 1992a. *Conditional Logics for Default Reasoning and Belief Revision.* Ph.D. Dissertation, Department of Computer Science, University of Toronto.

Boutilier, Craig 1992b. What is a default priority? In *Canadian Conference on AI*, Vancouver, B.C.

Brewka, G. 1991. Cumulative default logic: In defense of nonmonotonic inference rules. *Artificial Intelligence* 50(2):183–205.

Brewka, G. 1993. Adding priorities and specificity to default logic. Manuscript.

Delgrande, J. and Schaub, T. 1993. On using system Z to generate prioritised default theories. Technical report, IRISA. Forthcoming.

Delgrande, James P.; Jackson, W. Ken; and Schaub, Torsten 1994. Alternative approaches to default logic. *Artificial Intelligence.* to appear (accepted Feb. 1993).

Delgrande, J.P. 1987. A first-order conditional logic for prototypical properties. *Artificial Intelligence* 33(1):105–130.

Delgrande, J.P. 1988. An approach to default reasoning based on a first-order conditional logic: Revised report. *Artificial Intelligence* 36(1):63–90.

Etherington, D.W. and Reiter, R. 1983. On inheritance hierarchies with exceptions. In *Proc. AAAI-83.* 104–108.

Geffner, Hector and Pearl, Judea 1992. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence* 53(2-3):209–244.

Goldszmidt, Moisés and Pearl, Judea 1990. On the relation between rational closure and system Z. In *Third International Workshop on Nonmonotonic Reasoning*, South Lake Tahoe. 130–140.

Konolige, K. 1988. On the relation between default theories and autoepistemic logic. *Artificial Intelligence* 35(2):343–382.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44(1-2).

Lukaszewicz, W. 1988. Considerations on default logic: An alternative approach. *Computational Intelligence* 4(1):1–16.

McCarthy, J. 1980. Circumscription – a form of nonmonotonic reasoning. *Artificial Intelligence* 13:27–39.

Moore, R.C. 1985. Semantical considerations on nonmonotonic logic. *Artificial Intelligence* 25:75–94.

Pearl, J. 1989. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proc. KR-89*, Toronto. Morgan Kaufman. 505–516.

Pearl, J. 1990. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proc. of the Third Conference on Theoretical Aspects of Reasoning About Knowledge*, Pacific Grove, Ca. 121–135.

Poole, D. 1985. On the comparison of theories: Preferring the most specific explanation. In *Proc. IJCAI-85.* 144–147.

Poole, D.L. 1988. A logical framework for default reasoning. *Artificial Intelligence* 36(1):27–48.

Reiter, R. and Criscuolo, G. 1981. On interacting defaults. In *Proc. IJCAI-81*, Vancouver, B.C. 270–276.

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

Zhang, A. and Marek, W. 1990. On the classification and existence of structures in default logic. *Fundamenta Informaticae* 8(4):485–499.