Compiling Specificity into Approaches to Nonmonotonic Reasoning

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

Torsten H. Schaub
LERIA
Faculté des Sciences, Université d'Angers
2, boulevard Lavoisier
F-49045 Angers Cedex 01
torsten.schaub@univ-angers.fr

Abstract

We present a general approach for introducing specificity information into nonmonotonic theories. Historically, many approaches to nonmonotonic reasoning, including default logic, circumscription, and autoepistemic logic, do not provide an account of specificity, and so fail to enforce specificity among default sentences. In our approach, a default theory is initially given as a set of strict and defeasible rules. By making use of a theory of default conditionals, here given by System **Z**, we isolate minimal sets of defaults with specificity conflicts. From the specificity information intrinsic in these sets, a default theory in a target language is specified. For default logic the end result is a semi-normal default theory; in circumscription the end result is a set of abnormality propositions that, when circumscribed, yield a theory in which specificity information is appropriately handled. We mainly deal with default logic and circumscription although we also consider autoepistemic logic, Theorist, and variants of default logic and circumscription. This approach differs from previous work in that specificity information is obtained from information intrinsic in a set of conditionals, rather than assumed to exist a priori. Moreover, we deal with the "standard" version of, for example default logic and circumscription, and do not rely on prioritised versions, as

do other approaches. The approach is both uniform and general, so the choice of the ultimate target language has little effect on the overall approach.

1   Introduction

A general problem in many approaches to nonmonotonic reasoning is that they do not enforce *specificity* relations among default assertions as part of their basic machinery. Consider for example where birds fly, birds have wings, penguins must be birds, and penguins don't fly. We can write this as:

$$B \rightarrow F, \; B \rightarrow W, \; P \Rightarrow B, \; P \rightarrow \neg F. \tag{1}$$

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.

Default logic [Reiter, 1980], circumscription [McCarthy, 1980], and autoepistemic logic [Moore, 1985] are examples of approaches that do not take specificity information into account. For example, in the naïve 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 in which F is true. One is required to use so-called semi-normal defaults[1] to eliminate the second extension. However, it is up to the user to hand-code how specificity is dealt with. [Reiter and Criscuolo, 1981], for example, gives a partial list of ways of transforming default theories so that unwanted extensions arising from specific "interactions" are eliminated.

There are more recent approaches to nonmonotonic reasoning, based generally on intuitions from probability theory or conditional logics, that deal with specificity in a very natural way. Moreover, in the past few years there has been some consensus as to what should constitute a basic conditional system. 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, a green bird cannot be assumed to fly by default (since it is conceivable

---

[1]See Section 4.1 for a definition of semi-normal defaults and the way they deal with unwanted extensions.

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.

Our approach is to use the specificity information determined by a conditional system to generate a default theory in a nonmonotonic reasoning system, such as default logic, so that specificity is appropriately handled in the latter approach. Hence we address two related but essentially independent questions:

1. How can a conditional system be used to isolate interacting defaults with differing specificity?

2. How can this information be uniformly incorporated in a theory expressed in a nonmonotonic reasoning system where specificity is not directly addressed?

For the first part, we consider System **Z** [Pearl, 1990] as an example of a conditional system of defeasible reasoning. For the second part, we consider first *consistency-based* approaches, as exemplified by default logic [Reiter, 1980]; subsequently we consider variants of default logic and other related approaches. Second we consider *minimization-based* approaches, as exemplified by circumscription [McCarthy, 1980], and again variants and related systems.

We begin with a background theory made up of a set of strict rules $R_N = \{r \mid \alpha_r \Rightarrow \beta_r\}$ assumed to be true in every setting, together with a set of defeasible rules $R_D = \{r \mid \alpha_r \rightarrow \beta_r\}$, 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; 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 Section 3.1), but rather appeal to the *techniques* of this approach to isolate conflicting subsets of defaults. In a second step, we use the derived specificity information to produce, for instance, a set of default rules in default logic, or a classical theory that can be circumscribed, in such a way that specificity is suitably handled. The framework described then is a general approach to "compiling" default theories, using a conditional approach to determine specificity conflicts, into an approach to nonmonotonic reasoning where specificity is not "automatically" handled.

This framework offers several advantages over earlier work. First, it is more general, in that it is applicable to broad classes of systems, rather than to a specific system. In addition, within a specific system, the class of specificity conflicts

3

handled is broader than previous work, addressing for example the situation of a set of less-specific defaults with a more-specific default. Second, specificity information is obtained by appeal to an extant theory of defaults, and not simply some external user-specified ordering of defaults. So the present approach provides a justification for these modifications. Third, specificity is added to default logic (or autoepistemic logic, circumscription, etc.) without changing the machinery of default logic. That is, the resultant default theory is a theory *in* default logic, and not for example a set of ordered default rules requiring modifications to default logic. Hence we effectively remain within the original formalism, and so can take advantage of previous work (including implementations) concerning these approaches. In addition, we prove that specificity conflicts are indeed resolved in a general fashion, leaving unchanged other conflicts (as are found for example in a "Nixon diamond").

In the next subsection we briefly cover background material, while Section 2 introduces our approach. Section 3 shows how specificity conflicts are determined. Section 4 shows how specificity is compiled into consistency-based approaches to nonmonotonic reasoning, while Section 5 does the same for minimization-based approaches. Section 6 gives a brief summary. Portions of this work appeared earlier in [Delgrande and Schaub, 1994a, Delgrande and Schaub, 1994b].

1.1   Default Theories

Knowledge about the world, given in a knowledge base Δ, is assumed to be divided into two sets:

R : Background knowledge, or facts or rules which are assumed to be applicable in every domain.

W : Contingent knowledge, or facts which are true in the case under consideration and which may vary from case to case.

This is essentially the difference between necessary and contingent knowledge in modal logics [Hughes and Cresswell, 1968], or between probabilistic knowledge and conditioning knowledge in probabilistic reasoning systems [Pearl, 1989]. Background knowledge in turn consists of two sets:

R_D: Default implications, or rules that are usually true but allow exceptions.

R_N: Necessary implications, or rules which must be true in any setting.

This division is found in the various *conditional* approaches to default reasoning in Artificial Intelligence, such as [Delgrande, 1987, Geffner and Pearl, 1992,

4

Boutilier, 1992a, Goldszmidt, 1992] and less directly in [Kraus *et al.*, 1990].[2] The background knowledge provides a *generic world description*. Elements of W are formulas of classical logic. Elements of R_D are formulas of the form $\alpha \rightarrow \beta$ while elements of R_N are formulas of the form $\alpha \Rightarrow \beta$, where α and β are propositional formulas. The expression of elements of R_N as rules is a convenience only since an arbitrary strict formula α can be expressed by $\top \Rightarrow \alpha$. Note too that we reserve $\supset$ for classical (material) implication.

So a knowledge base is of the form $\Delta = \langle\langle R_D, R_N \rangle, W\rangle$, where $\langle R_D, R_N \rangle$ represents generic world knowledge and W represents case-specific knowledge. Our initial example is represented as:

$$R = \langle\{B \rightarrow F,\ B \rightarrow W,\ P \rightarrow \neg F\}, \{P \Rightarrow B\}\rangle. \tag{2}$$

It should be obvious how these sets would be mapped into a particular approach to nonmonotonic reasoning. In default logic, for example, elements of R_D would be mapped into default rules, and everything else would be considered as world knowledge; in circumscription, elements of R_D would be mapped to implications with ab propositions, while again everything else would be considered as world knowledge. However, as the previous section pointed out, the "obvious" mappings are problematic in that specificity is not properly handled. The purpose of this paper then might be seen as developing a general, provably correct "compilation" scheme to address specificity for those approaches to nonmonotonic reasoning that, as part of their formalism, do not.

1.2 Related Work

Arguably, specificity per se was first specifically addressed in default reasoning in [Poole, 1985], although it has of course appeared earlier. As mentioned, we could have used a conditional system other than System **Z** in our approach; however, System **Z** is particularly straightforwardly describable. Some approaches though are too weak to be useful here. For example conditional entailment [Geffner and Pearl, 1992] does not support full inheritance reasoning, in that from $\{A \rightarrow B,\ B \rightarrow C,\ C \rightarrow D,\ A \rightarrow \neg D\}$ we cannot conclude C by default from A. [Delgrande, 1988] is unsatisfactory since it gives a syntactic, albeit general, approach in a conditional logic.

Conditional approaches are founded, one way or another, on notions of *preference* or *normality*. Conditional logics for default conditionals [Delgrande, 1987, Boutilier, 1992a], for example, are modal logics [Hughes and Cresswell, 1968, Chellas, 1980]

---

[2]The roots of such approaches however extend at least as far back as [Adams, 1975, Lewis, 1973, Stalnaker, 1968].

where we can view possible worlds as being ordered by a metric of normality or un-exceptionalness. A default conditional $\alpha \to \beta$ is true, roughly, if in the least worlds where α is true, β is true also. For our example (1), at the least worlds where B is true, F is true also; at the least worlds where P is true, $\neg F$ is true. Since we also have that $P \Rightarrow B$ (we could as easily have $P \to B$), it can be seen that the least P worlds must be more exceptional than (or less normal than) the least B worlds. If we now say that β follows as a default inference from α in default theory R just when β is true in the least α worlds, we obtain a form of default inference in which specificity is obtained and, for example, penguins normally don't fly whereas birds do.

These approaches are quite weak: since it is conceivable that green birds do not fly (i.e. there are models where in the least green-bird worlds these birds do not fly), it does not follow by default that a green bird flies, even though a bird does. While various approaches have been proposed to strengthen such basic systems, including rational closure [Kraus *et al.*, 1990], System **Z**, CO^* [Boutilier, 1992a], possibilistic entailment [Benferhat *et al.*, 1992], and conditional objects [Dubois and Prade, 1991] none is entirely satisfactory. In this regard, System **Z** is examined as an exemplar of these approaches in Section 3.1.

In default logic, [Reiter and Criscuolo, 1981] considers various patterns of specificity in interacting defaults, and describes how specificity may be obtained via semi-normal defaults. However these patterns are just commonly-occurring configurations of defaults, and there is no notion of this being a complete characterisation. This work 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 Section 3) 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, and this network is compiled into a default theory in default logic (see Section 4.3).

More recent work develops priority orderings on default theories, including [Boutilier, 1992b, Baader and Hollunder, 1993a, Brewka, 1993]. However these approaches obtain specificity by requiring modifications to how default logic is used. In contrast, we describe transformations that yield classical default logic theories. Since these approaches are described in Section 4, they are introduced only briefly here. [Boutilier, 1992b] uses the correspondence between a conditional $\alpha_r \to \beta_r$ of System **Z** and defaults of the form $\frac{: \alpha_r \supset \beta_r}{\alpha_r \supset \beta_r}$ to produce partitioned sets of prerequisite-free normal default rules. One reasons in this approach by applying the rules in the highest set, and working down. [Baader and Hollunder, 1993a] addresses specificity in terminological reasoners. 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 default logic.

[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 default logic are used; however, for inferencing all consistent strict total orders of defaults must be considered.

Similar remarks apply to circumscription, and to other related approaches. Circumscription was introduced in [McCarthy, 1980], and the use of ab predicates in prioritised, parameterised circumscription to address specificity was addressed in [McCarthy, 1986] and [Lifschitz, 1985]. [Grosof, 1991] extended prioritised circumscription to deal with partial orders. These approaches are discussed more fully in Section 5.3.

Lastly there are *direct* or *path-based* approaches to nonmonotonic inheritance, as expressed using inheritance networks [Horty, 1994]. It is difficult to compare such approaches with our own for two reasons. First, inheritance networks are concerned, broadly, with general notions having to do with arguments or nonmonotonic inheritance. Our interests are narrower, being limited to specificity. Second, the account of meaning for such networks is most often given in terms of paths in the network, and so tend not to rely on more standard model-theoretic notions. Nonetheless, in Section 4.3 we compare our approach with probably the best-know translation of an inheritance network, where the network is translated into a theory in default logic, that of [Etherington and Reiter, 1983].

2   Overview of the Approach

There are two major steps in the approach. First, given a default theory expressed as a generic world description, we locate default rules that conflict and have differing specificity; this is accomplished by using (part of) the mechanism of System **Z**. So for our initial example (1) it is clear that the defaults $B \rightarrow F$ and $P \rightarrow \neg F$ conflict and that the second default is more specific than the first. Secondly, we compile the default theory into a nonmonotonic reasoning system such as default logic or circumscription, so that if both defaults are potentially applicable—say, B and P are true—then only the second default is applied. In outline, this is carried out as follows.

In System **Z** defaults are partitioned into sets R_0, R_1, ..., where, roughly, the defaults in a lower-ranked partition are less specific than those in a higher-ranked partition. The resulting partition is called a **Z**-*ordering*. For our initial example, treating the strict implication as a default for the moment, we would obtain the partition:

$$
\begin{aligned}
R_0 &= \{B \rightarrow F,\ B \rightarrow W\} \\
R_1 &= \{P \rightarrow B,\ P \rightarrow \neg F\}.
\end{aligned}
$$

The key point in determining the partition is that, if we treat $\rightarrow$ as classical implication, then for $\alpha \rightarrow \beta \in R_i$ we have that

$$\{\alpha \wedge \beta\} \cup R_i \cup R_{i+1} \cup \ldots$$

is satisfiable, whereas

$$\{\alpha \wedge \beta\} \cup R_{i-1} \cup R_i \cup R_{i+1} \cup \ldots$$

is unsatisfiable. So the **Z**-ordering provides specificity information; however, we do not use the full **Z**-ordering since it may introduce unwanted specificities (Section 3.1). Rather we determine minimal sets of rules that conflict, and use these sets to sort out specificity information. In the above example, $\{B \rightarrow F, P \rightarrow B, P \rightarrow \neg F\}$ would be such a set, since if we delete any of the three defaults we would have a set with no conflict or with no difference in specificity.

There are numerous issues that need to be confronted, even with relatively simple theories. Consider the following extended example, already expressed as a **Z**-ordering; we will make reference to this example throughout the paper. For simplicity we have expressed all rules as default rules.

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

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 (with respect to set inclusion) sets of rules that differ in specificity and that conflict; this will be the minimal set of rules having a non-trivial **Z**-ordering. In our example these consist of:

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

Any such set is called a *minimal conflicting set* of rules. For any such set, if all the rules are jointly applicable then one way or another there will be a conflict.[3] Note

---

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

that both of these notions are crucial. If we have a conflict without a specificity difference, for example with the defaults,

$$Q \rightarrow P, \ R \rightarrow \neg P$$

then given $Q \wedge R$ there is no reason to apply one rule over another. If we have a specificity difference without a conflict, say birds fly and tropical birds are colourful:

$$B \rightarrow F, \ B \wedge T \rightarrow C$$

Then given $B \wedge T$ there is again no reason to not apply both defaults.

We show below that the **Z**-ordering of a each such set C consists of a binary partition (C_0, C_1) of rules. 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. For example, for the minimal conflicting set C^1 we obtain:

$$\begin{aligned} C_0^1 &= \{B \rightarrow F\} & (5)\\ C_1^1 &= \{P \rightarrow B, \ P \rightarrow \neg F\}. & (6) \end{aligned}$$

There are now two important issues that need to be addressed:

1. What rules should be selected as candidates to be blocked, using minimal conflicting sets?

2. How can the application of a rule be blocked in the target nonmonotonic formalism?

For the first question, consider where we have a chain of rules, and where transitivity is explicitly blocked, as in the minimal conflicting set C^4 above. We have the **Z**-ordering:

$$\begin{aligned} C_0^4 &= \{An \rightarrow WB, B \rightarrow An\} & (7)\\ C_1^4 &= \{Pt \rightarrow B, Pt \rightarrow \neg WB\}. \end{aligned}$$

Intuitively An is less specific than Pt. 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 $Pt \rightarrow B$ is falsified. So we want more specific rules to be applicable over less specific conflicting rules, independently of the other rules in the minimal conflicting set. We do this by locating those rules whose joint applicability would lead to an inconsistency. In our example, this consists of $An \rightarrow WB$, and $Pt \rightarrow \neg WB$

9

(since $(An \wedge WB) \wedge (Pt \wedge \neg WB)$ is inconsistent). Since $An \rightarrow WB \in C_0^4$ and $Pt \rightarrow \neg WB \in C_1^4$, the rules have differing specificity. The rules selected in this way from C_0^4 and C_1^4 are called the *minimal conflicting rules* and *maximal conflicting rules* respectively. The minimal conflicting rules constitute the candidates to be blocked. This selection criterion has the important property that it is *context independent*, 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'.

The second question, (*"How can the application of a rule be blocked?"*) depends on the target nonmonotonic formalism. However we argue that our approach is broadly applicable to nonmonotonic formalisms that do not, in and of themselves, address specificity issues. In Section 4 we deal with the major consistency-based formalisms; Section 5 addresses minimization-based formalisms. For default logic[4] for example, we have the following translation of rules. The default theory corresponding to our default rules R_D consists of normal defaults, except for those defaults representing minimal conflicting rules, which are 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 minimal conflicting set cannot be applicable.

Consider the set C_0^4 in (7), along with its minimal conflicting rule $An \rightarrow WB$. We replace $B \rightarrow An$, $Pt \rightarrow B$, $Pt \rightarrow \neg WB$ with

$$\frac{B : An}{An}, \quad \frac{Pt : B}{B}, \quad \frac{Pt : \neg WB}{\neg WB}$$

respectively. For $An \rightarrow WB$, we replace it with

$$\frac{An : WB \wedge (Pt \supset \neg WB)}{WB}, \quad \text{which can be simplified to} \quad \frac{An : WB \wedge \neg Pt}{WB}.$$

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 minimal conflicting sets C^0 to C^4:

$$
\begin{array}{lll}
C^0 : & \dfrac{An : \neg Fe \wedge (B \supset Fe)}{\neg Fe} & \text{or} \quad \dfrac{An : \neg Fe \wedge \neg B}{\neg Fe} \\[2ex]
C^1 + C^2 : & \dfrac{B : F \wedge (P \supset \neg F) \wedge (E \supset \neg F)}{F} & \text{or} \quad \dfrac{B : F \wedge \neg P \wedge \neg E}{F} \\[2ex]
C^3 : & \dfrac{B : Fe \wedge (Pt \supset \neg Fe)}{Fe} & \text{or} \quad \dfrac{B : Fe \wedge \neg Pt}{Fe} \\[2ex]
C^4 : & \dfrac{An : WB \wedge (Pt \supset \neg WB)}{WB} & \text{or} \quad \dfrac{An : WB \wedge \neg Pt}{WB}
\end{array}
$$

---

[4]A formal introduction to default logic is given in Section 4.1.

The conditional $B \to F$ occurs in C^1 and C^2 as a minimal conflicting rule. In this case we have two minimal conflicting sets 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 where we have a minimal conflicting set of defaults C with a single minimal conflicting rule $\alpha_0 \to \beta_0$ and a single maximal conflicting rule $\alpha_1 \to \beta_1$. If we can prove that α_0 (and so in default logic can prove the antecedent of the conditional), then β_0 may be a default conclusion, provided that no more specific rule applies. But what should constitute the justification? Clearly, that β_0 is consistent and that more specific, conflicting conditionals not be applicable. Now, in our setting, $\alpha_0 \to \beta_0$ is such that $\{\alpha_0 \wedge \beta_0\}$ is satisfiable, but for the conditional $\alpha_1 \to \beta_1$, $\{\alpha_0 \wedge \beta_0\} \cup \{\alpha_1 \wedge \beta_1\}$ 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. Hence we add these more specific conditionals as part of the justification.

We show too that this approach is applicable to general default theories and not, as the preceding examples might indicate, just simple chains of defaults. Consider the following example, in which we have two less-specific default rules, a situation frequently found in multiple inheritance networks. We have the **Z**-ordering:

$$R_0 \quad = \quad \{A \to \neg B, C \to \neg D\} \tag{8}$$
$$R_1 \quad = \quad \{A \wedge C \to B \vee D\} \tag{9}$$

In this case we would want to ensure that if the default in R_1 were applicable, then at most one default in R_0 can be applied. 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 in detail following the presentation of the formal details.

3 Determining Specificity Conflicts

3.1 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 in lower-ranked sets are *compatible* with those in higher-ranked sets, whereas rules in higher-ranked sets *conflict* in some fashion with rules in lower-ranked sets. [Pearl, 1990] deals only with default rules, whereas the extension described in [Goldszmidt, 1992] deals with default and strict rules. For our use of System **Z**

we do not need to distinguish default and strict rules, [5] and so we describe the original approach. We assume that we begin with a set of defeasible conditionals $R = \{r \mid \alpha_r \to \beta_r\}$ where each α_r and β_r are propositional formulas over a finite alphabet. A central notion is that of *toleration*:

Definition 1 *Let R be a set of defeasible conditionals.*
  A rule $\alpha \to \beta$ is tolerated by R iff $\{\alpha \wedge \beta\} \cup \{\alpha_r \supset \beta_r \mid r \in R\}$ is satisfiable.

Note that this definition treats the connective $\to$ as $\supset$.
  We assume in what follows that R is **Z**-*consistent*,[6] i.e. for every non-empty $R' \subseteq R$, some $r' \in R'$ is tolerated by R'. Using this notion of tolerance, a **Z**-*ordering* on the rules in R is defined:

1. 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 in R 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, or its **Z**-ordering, respectively, 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 interpretation M of R is given a **Z**-rank, $\mathbf{Z}(M)$, according to the highest ranked rule in R it falsifies:

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

When interpreting all rules in our example (1) as defeasible, we obtain the following **Z**-ordering:

$$
\begin{aligned}
R_0 &= \{B \to F,\ B \to W\} \\
R_1 &= \{P \to \neg F,\ P \to B\}.
\end{aligned}
\tag{10}
$$

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

---

[5]Essentially we use System **Z** to isolate conflicting rules, independent of whether they are strict or default. This distinction is important for us only when deciding on what rules to block.
  [6][Pearl, 1990] uses the term *consistent*.

as the lowest **Z**-rank of all models satisfying φ: $\mathbf{Z}(\varphi) = min\{\mathbf{Z}(M) \mid M \models \varphi\}$.[7] Finally we can define a form of default entailment, called *1-entailment*, as follows: A formula φ is said to 1-entail ϕ in the context R, written $\varphi \vdash_1 \phi$, iff $\mathbf{Z}(\varphi \wedge \phi) < \mathbf{Z}(\varphi \wedge \neg\phi)$. In the terminology of Section 1.1, the background theory R determines a **Z**-ordering, and α follows from our contingent knowledge W iff $W \vdash_1 \alpha$.

This gives a form of default inference that 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 default logic, we cannot infer that penguins fly, i.e. $P \not\vdash_1 F$. Some irrelevant facts are 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, if we add to the above example the default that large animals are calm we get the **Z**-ordering:

$$R_0 = \{B \rightarrow F,\ B \rightarrow W,\ L \rightarrow C\} \tag{11}$$

$$R_1 = \{P \rightarrow \neg F,\ P \rightarrow B\}. \tag{12}$$

Intuitively $L \rightarrow C$ is irrelevant to the other defaults, yet one obtains the default conclusion that large animals aren't normally penguins 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. [Pearl, 1990] notes that preferential entailment [Lehmann, 1989] is equivalent to the more basic notion of 0-entailment (also ϵ-entailment [Pearl, 1988] or p-entailment [Adams, 1975]), 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.

3.1.1 Why System Z?

The previous subsection described System **Z**, which we use to isolate minimal sets of rules (strict and default) that conflict with respect to specificity. The natural questions arise, why choose System **Z** when, as indicated previously, it is not unproblematic? And, are there alternatives to the choice of System **Z**?

---

[7]If there is no model satisfying ϕ we set the rank of ϕ as ∞.

First of all, we do not use System **Z** per se, but rather the notion of *tolerance*; this we use to isolate minimal sets of rules with a nontrivial partition. In such (minimal) sets the problems of unwanted specificities do not arise (since there are no "irrelevant" rules). Moreover, we are unconcerned about lack of property inheritance since we obtain such inheritance in the target language, whether it be default logic, circumscription, or some other.

In the second case, while there are approaches that could be used in place of System **Z**, System **Z** (or the part that we use) is certainly the simplest. For those familiar with conditional logics (or related approaches) we note that a system corresponding to the conservative core is too weak for our purposes. In particular, such a system allows the conditionals

$$\{\alpha \to \gamma, \ \neg(\alpha \wedge \beta \to \gamma), \ \neg(\alpha \wedge \neg\beta \to \gamma)\}$$

to be simultaneously and non-trivially satisfied. For a logic of defaults, this appears unreasonable: if γ follows by default from α, then it would seem that it should also follow from either $\alpha \wedge \beta$ or $\alpha \wedge \neg\beta$. Arguably the weakest logic in which $\alpha \to \gamma \supset ((\alpha \wedge \beta \to \gamma) \vee (\alpha \wedge \neg\beta \to \gamma))$ is a theorem, is N [Delgrande, 1987]. If we do not consider negated conditionals, then this is equivalent to VTA [Lewis, 1973] or CO [Boutilier, 1992a], and is the conditional equivalent of S4.3 [Hughes and Cresswell, 1968]. While these latter systems could be used as a basis from which specificity information could be determined, System **Z** is markedly easier to describe than these other approaches; moreover determining 1-entailment is efficient (disregarding consistency tests).

3.2 Minimal Conflicting Sets

We consider **Z**-consistent generic world descriptions $R = \langle R_D, R_N \rangle$ where the antecedents and consequents of rules in R are propositional formulas over a finite alphabet. For simplicity, we sometimes identify R with $R_D \cup R_N$. For **Z**-orderings of subsets of R, we treat the connective $\Rightarrow$ as $\to$ (that is, we do not distinguish strict and default rules in an ordering). We denote the set of classical implications corresponding to a set R of strict and/or defeasible rules by R^*. That is,

$$R^* = \{\alpha \supset \beta \mid \alpha \to \beta \in R_D\} \cup \{\alpha \supset \beta \mid \alpha \Rightarrow \beta \in R_N\}.$$

Moreover, we define

$$\begin{aligned}
\mathit{Prereq}(R) &= \{\alpha \mid \alpha \to \beta \in R_D\} \cup \{\alpha \mid \alpha \Rightarrow \beta \in R_N\} \quad \text{and} \\
\mathit{Conseq}(R) &= \{\beta \mid \alpha \to \beta \in R_D\} \cup \{\beta \mid \alpha \Rightarrow \beta \in R_N\}.
\end{aligned}$$

The set of minimal conflicting sets of a set of rules R represents conflicts among the rules in R due to disparate specificity. Each minimal conflicting set is a minimal set of conditionals having a non-trivial **Z**-ordering.

Definition 2 *Let $R = \langle R_D, R_N \rangle$ be a generic world description. A set of rules $C \subseteq R$ is a minimal conflicting set in R iff C has a non-trivial $\mathbf{Z}$-ordering and any $C' \subset C$ has a trivial $\mathbf{Z}$-ordering.*

That is, the rules in C make up a nontrivial $\mathbf{Z}$-ordering and they form a least set for which a nontrivial ordering is obtained. A minimal conflicting set then constitutes a minimal theory in which there is a specificity conflict. Observe that adding new rules to R cannot alter or destroy any existing minimal conflicting sets. That is, for default theories R and R', where $C \subseteq R \subseteq R'$, we have that if C is a minimal conflicting set in R then C is a minimal conflicting set in R'. This property is of great practical relevance since it allows an incremental computation of minimal conflicting sets, even in evolving knowledge bases.

The next theorem shows that any minimal conflicting set has a binary partition:

Theorem 1 *Let C be a minimal conflicting set in some generic world description $\langle R_D, R_N \rangle$. Then, we have that the $\mathbf{Z}$-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 minimal conflicting set entails the negations of the antecedents of the higher-level rules:

Theorem 2 *Let C be a minimal conflicting set in R. If $\alpha \to \beta \in C_1$ then $C^* \models \neg\alpha$.*

Hence, given our initial generic world description in (2),

$$R = \langle\ \{B \to F,\ B \to W,\ P \to \neg F\}\ ,\ \{P \Rightarrow B\}\ \rangle$$

there is one minimal conflicting set

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

As shown in (5/6), the first rule constitutes C_0 and the last two C_1 in the $\mathbf{Z}$-ordering of C (in fact, the last rule provides rather necessary linking knowledge, as explicated in (13). If we discard the necessary knowledge provided by $P \Rightarrow B$, the set $\{B \to F,\ P \to \neg F\}$ is not a minimal conflicting set since alone it has a trivial $\mathbf{Z}$-ordering. Replacing $P \Rightarrow B$ by $P \to B$ yields obviously the same minimal conflicting set. It is easy to see that $C^* \models \neg P$.

Intuitively, a minimal conflicting set 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 general formal frame for these sets:

Definition 3 *Let R be a generic world description and let C be a minimal conflicting set in R. We define $max(C)$, $min(C)$, and $inf(C)$ to be non-empty subsets of C such that*

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

We observe that *min*, *max*, and *inf* are exclusive subsets of C such that $C = min(C) \cup inf(C) \cup max(C)$. 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 $max(C)$. However following 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" in $inf(C)$:

Theorem 3 *Let C be a minimal conflicting set in a generic world description $\langle R_D, R_N \rangle$. Then, we have:*

$$
inf(C)^* \cup max(C)^* \models \textbf{\textit{Prereq}}(max(C)) \supset \textbf{\textit{Prereq}}(min(C)).
$$

In fact, $inf(C)^* \cup max(C)^*$ constitutes the weakest condition under which the above entailment holds. Note that omitting $max(C)$ would eliminate rules that may belong to $max(C)$, yet provide "inferential relations". This is the case for the rule $P \Rightarrow B$ in (5/6): $P \Rightarrow B$ is in C_1 and so is a potential candidate for $max(C)$, even though this choice is not a reasonable one (since of course, $P \Rightarrow B$ should be a part of $inf(C)$). The same applies of course to a defeasible rule like $P \rightarrow B$.

The next theorem shows that the converse of the previous does not hold in general.

Theorem 4 *Let C be a minimal conflicting set in a generic world description $\langle R_D, R_N \rangle$. Then, for any set of rules R' such that $C \subseteq R'$ and any set of rules $R'' \subseteq min(C)$ such that $R' \cup \textbf{\textit{Prereq}}(R'')$ is satisfiable, we have:*

$$
(R')^* \not\models \textbf{\textit{Prereq}}(R'') \supset \textbf{\textit{Prereq}}(max(C)).
$$

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, $C \cup \textbf{\textit{Prereq}}(min(C))$ and so $R' \cup \textbf{\textit{Prereq}}(min(C))$ is inconsistent. This is, for instance, the case in (8/9). In fact, $(R')^*$ is the strongest condition under which the above theorem holds.

Finally, we demonstrate that these rules are indeed conflicting.

Theorem 5 *Let C be a minimal conflicting set in a generic world description $\langle R_D, R_N \rangle$. Then, for any $\alpha \rightarrow \beta \in max(C)$, we have:*

$$inf(C)^* \cup \{\alpha\} \models \neg(Conseq(min(C)) \wedge Conseq(max(C))).$$

As above, $inf(C)^* \cup \{\alpha\}$ is the weakest condition under which the last entailment holds. In all, the last three theorems demonstrate that the general framework given for minimal conflicting sets (already) provides a very expressive way of isolating rule conflicts due to their specificity.

In the worst case the number of minimal conflicting sets grows exponentially with the size of a default theory. This is an artifact of the problem in general, rather than the specific approach at hand – there may simply be an exponential number of ways in which a set of defaults conflict.

Theorem 6 *There exist generic world descriptions R of size n such that the number of minimal conflicting sets is of size $O(2^n)$.*

Consider for example the class of default theories where we have

$$
\begin{aligned}
\alpha \rightarrow \beta_{i,1} \quad &\text{for} \quad i \in \{1, 2\} \\
\beta_{i,j} \rightarrow \beta_{i',j+1} \quad &\text{for} \quad i, i' \in \{1, 2\} \text{ and } 1 \leq j < n \\
\beta_{i,n} \rightarrow \gamma \quad &\text{for} \quad i \in \{1, 2\}
\end{aligned}
$$

For a given n there are clearly 2^n "inferential paths" between α and γ. If we add the default $\alpha \rightarrow \neg\gamma$, then clearly for given n there are 2^n minimal conflicting sets. While this characterises the worst case, in general we might expect the number of minimal conflicting sets to be more manageable. For example, in an inheritance hierarchy where a different "exception" type accounts for each level in the hierarchy, we would have a set of minimal conflicting sets that is linear in the size of the hierarchy.

3.3 Specific Minimal and Maximal Conflicting Rules

A minimal conflicting set $C = (C_0, C_1)$, is a minimal set of rules that contains a specificity conflict. However we need to isolate a minimal subset $C_0' \subseteq C_0$ whose application would conflict with a minimal subset of rules in $C_1' \subseteq C_1$. We do this in the following definition of a *conflicting core* of a minimal conflicting set:

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

    1. $min(C) \subseteq C_0 \cap R_D$,

2. $max(C) \subseteq C_1 \cap R_D$,

3. $\{\alpha_r \wedge \beta_r \mid r \in max(C) \cup min(C)\} \models \bot$,

provided that $min(C)$ and $max(C)$ are non-empty.

This definition specialises the general setting of Definition 3. So, $\alpha_r \rightarrow \beta_r$ is in $min(C)$ if its application conflicts with that of a rule (or rules) in C_1. By isolating the actually conflicting rules in C_0 and C_1, a conflicting core imposes a strict structure on a given minimal conflicting set.

It is only with Definition 4 that we distinguish default from strict rules, in that we eliminate members of R_D from $min(C)$ and $max(C)$. In Section 4.2 we show that this is a convenience only, in that if we included members of R_D in $min(C)$ and $max(C)$ we would simply introduce redundant elements into our default theory. However, consider informally the effect of strict rules in Definition 4. For example, the rules $B \rightarrow F$ and $P \rightarrow \neg F$ yield the conflicting core $(\{B \rightarrow F\}, \{P \rightarrow \neg F\})$ in our example (1). This induces the following structure on the minimal conflicting set C given in (5/6):

$$min(C) = \{B \rightarrow F\} \qquad max(C) = \{P \rightarrow \neg F\} \qquad inf(C) = \{P \Rightarrow B\} \tag{13}$$

Observe that we obtain the same conflicting core for C^1 in (3), where $P \Rightarrow B$ is treated defeasibly by means of $P \rightarrow B$. That is, the applicability of $P \rightarrow B$ (or $P \Rightarrow B$) is irrelevant to the conflict between $B \rightarrow F$ and $P \rightarrow \neg F$, and so can be applied independently of these last two defaults. However things are quite different if we replace $B \rightarrow F$ or $P \rightarrow \neg F$ by their strict counterpart. In the theory

$$\{B \rightarrow F, P \rightarrow B, P \Rightarrow \neg F\}$$

for example, there is no conflict. If P is true or if $P \wedge B$ is true, then it logically follows that $\neg F$ is true, and so we cannot "apply" the default rule $B \rightarrow F$, regardless of the "target" formalism. On the other hand, in the theory

$$\{B \Rightarrow F, P \rightarrow B, P \rightarrow \neg F\}$$

we lose our specificity difference. If P is true, then application of $P \rightarrow B$ (regardless of how this is done) immediately blocks $P \rightarrow \neg F$ and application of $P \rightarrow \neg F$ immediately blocks $P \rightarrow B$.

In the extended example of Section 2 the conflicting cores are

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

respectively. As anticipated in Section 2, we thus obtain the following structure for minimal conflicting set C^4 (given as a **Z**-ordering in (7)):

$$min(C^4) = \{An \rightarrow WB\} \quad max(C^4) = \{Pt \rightarrow \neg WB\} \quad inf(C^4) = \{B \rightarrow An, \; Pt \rightarrow B\}$$

The remaining sets $max(C^i)$, $min(C^i)$, and $inf(C^i)$ are constructed in the obvious way.

For a complement consider the example given in (8/9), where the conflicting core contains two minimal and one maximal conflicting rules:

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

That is, $min(C) = \{A \rightarrow \neg B, C \rightarrow \neg D\}$, $max(C) = \{A \wedge C \rightarrow B \vee D\}$, and $inf(C) = \emptyset$.

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

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

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. In such a case, according to Definition 4, there is no conflicting core. We do have the following result however.

Theorem 7 *For minimal conflicting set C in a set of rules R, if $\{\alpha_r \wedge \beta_r \mid r \in min(C)\} \not\models \bot$ and $\{\alpha_r \wedge \beta_r \mid r \in max(C)\} \not\models \bot$ then C has a conflicting core.*

Note that while in "normal" cases a minimal conflicting set (apparently) has a unique conflicting core, this is not always the case. Consider the following minimal conflicting set:

$$
\begin{aligned}
C_0 &= \{\neg \alpha_1 \vee \neg \alpha_2 \rightarrow \beta_1\} \\
C_1 &= \{\alpha_1 \rightarrow \neg \beta_1 \wedge \beta_2, \; \alpha_2 \rightarrow \neg \beta_1 \wedge \neg \beta_2\}.
\end{aligned}
$$

We have two conflicting cores, since

$$
\begin{aligned}
\{\neg \alpha_1 \vee \neg \alpha_2, \beta_1\} \cup \{\alpha_1, \neg \beta_1 \wedge \quad \beta_2\} &\models \bot \quad \text{and} \\
\{\neg \alpha_1 \vee \neg \alpha_2, \beta_1\} \cup \{\alpha_2, \neg \beta_1 \wedge \neg \beta_2\} &\models \bot.
\end{aligned}
$$

This example is the only one that we have been able to construct in which there is a non-unique conflicting core. In the sequel, for simplicity we restrict our attention to minimal conflicting sets having a unique conflicting core. Non-unique conflicting cores are easily handled in Definition 6 by considering each minimal conflicting set/conflicting core pair separately.

19

4 Compiling Specificity into Consistency-Based Approaches

In the previous section, we described how to isolate minimal sets of rules that contain conflicting rules with differing specificity. We also showed how to isolate specific minimal and maximal conflicting rules. In this section, we use this information for specifying blocking conditions or, more generally, priorities among conflicting defaults in default logic.

There are two obvious approaches. First, we could determine a strict partial order on a set of rules R_D from the minimal conflicting sets in R. That is, for two rules $r, r' \in R_D$, we can define $r < r'$ iff $r \in min(C)$ and $r' \in max(C)$ for some minimal conflicting set C in R. In this way, $r < r'$ is interpreted as "r is less specific than r'". Then, one could interpret each rule $\alpha \to \beta$ in R_D 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. Such an approach has the disadvantage that it steps outside the machinery of default logic for computing extensions.

This motivates our primary approach, one that remains inside the framework of classical default logic, where we transform rules with specificity information into semi-normal default theories. After an introduction to default logic we develop this latter approach and explore its properties. Following this we compare our approach with that of related approaches in default logic. Lastly, we show how this approach can be applied to other consistency-based approaches to nonmonotonic reasoning.

4.1 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 $\omega \in E$. One of the simplest definitions of an extension, due to [Reiter, 1980], is the following:

Definition 5 *Let (D, W) be a default theory and let E be a set of formulas. Define $E_0 = W$ and for $i \geq 0$*

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

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

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

In terms of our specification of default theories, we assume that a default theory is given by $\langle\langle R_D, R_N\rangle, W\rangle$ whereas in default logic it is given by a pair (D', W'). The naïve translation for default logic is to identify the set of defeasible rules R_D with the set of normal default rules

$$\left\{ \frac{\alpha : \beta}{\beta} \ \Big| \ \alpha \rightarrow \beta \in R_D \right\},$$

while the strict rules in R_N are interpreted as material implications

$$R_N^* = \{\alpha \supset \beta \mid \alpha \Rightarrow \beta \in R_N\}.$$

In this way, a world description $\langle\langle R_D, R_N\rangle, W\rangle$ may be transformed into a default theory

$$\left(\left\{ \frac{\alpha : \beta}{\beta} \ \Big| \ \alpha \rightarrow \beta \in R_D \right\}, W \cup R_N^* \right).$$

Consider our example (1) along with the fact that P is true; this can be expressed as:

$$\left(\left\{ \frac{B : F}{F}, \ \frac{B : W}{W}, \ \frac{P : \neg F}{\neg F} \right\}, \{P\} \cup \{P \supset B\} \right). \tag{14}$$

We obtain two extensions: one in which $P, B, W,$ and F are true and another 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 solution, originally proposed in [Reiter and Criscuolo, 1981], is to establish a precedence among these two interacting defaults by adding the negation of 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 yields the desired result, a single extension containing $P, B, W,$ and $\neg F$.

4.2 Z-Default Theories

This section describes translation for producing a standard semi-normal default theory that provably maintains specificity. The transformation is succinctly defined:

Definition 6 *Let* $R = \langle R_D, R_N\rangle$ *be a generic world description and let* $(C^i)_{i \in I}$ *be the family of all minimal conflicting sets in R. For each* $r \in R_D$, *we define*

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

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

21

In what follows, we write $D_{R'} = \{\delta_r \mid r \in R'\}$ for any subset R' of R_D. Any default theory obtained according to the above transformation will be referred to as a **Z**-*default theory*.

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 minimal conflicting sets, 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 $r' \in max(C^i)$). These rules are put together in R_r. In this way, R_r contains all rules that conflict with r while being more specific than r. Hence, for the minimal conflicting rules we obtain semi-normal defaults; all other defaults are normal (since then $R_r = \emptyset$). So for any minimal conflicting set 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.

Consider our initial example in (1). There, we obtain a single minimal conflicting set C given in (5/6), having a unique conflicting core. As shown in (13), the latter induces the minimal and maximal conflicting rules: $min(C) = \{B \to F\}$ and $max(C) = \{P \to \neg F\}$. According to Definition 6, we obtain for the defeasible rules in (1):

$$R_{B \to F} = \{P \to \neg F\}, \ R_{B \to W} = \emptyset \ \text{and} \ R_{P \to \neg F} = \emptyset.$$

In turn, these sets induce the following default rules:

$$\delta_{B \to F} = \frac{B : F \wedge (P \supset \neg F)}{F}, \ \delta_{B \to W} = \frac{B : W}{W} \ \text{and} \ \delta_{P \to \neg F} = \frac{P : \neg F}{\neg F} \ .$$

The first rule can be simplified to $\frac{B : F \wedge \neg P}{F}$.

Given an entire world description $\langle\langle R_D, R_N \rangle, W \rangle$, we can apply Definition 6 in order to obtain **Z**-*default theory* $(D_R, W \cup R_N^*)$. Our initial example along with the contingent fact that P is true is then translated into the following **Z**-default theory:

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

As opposed to the naïve translation given in (14), this theory yields only the single, specificity-preserving extension, in which P, B, W, and $\neg F$ are true.

In the extended example of Section 2 the conflicting cores for (3) and (4) are

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

respectively. According to Definition 6, we get

$$R_{B \to F} = \{P \to \neg F, E \to \neg F\}, \ R_{P \to B} = \emptyset \ \text{and} \ R_{P \to \neg F} = \emptyset.$$

The first set expresses the fact that the rule $B \to F$ conflicts with the two more specific rules in $\{P \to \neg F, E \to \neg F\}$. 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.

These examples suggest that we might simply add the negation of the antecedent of the higher-level conflicting conditional. However this strategy does not work whenever a minimal conflicting set has more than one minimal conflicting rule. We defer the discussion to Section 4.3.

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

$$\frac{\alpha_r : \beta_r \wedge (\alpha_{r'} \supset \beta_{r'})}{\beta_r} \quad \text{and} \quad \frac{\alpha_{r'} : \beta_{r'}}{\beta_{r'}}.$$

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 8 is with respect to the general theory of minimal conflicting sets while Theorem 9 is with respect to the specific development involving conflicting cores.

Theorem 8 *Let $\langle \langle R_D, R_N \rangle, W \rangle$ be a world description with $R = \langle R_D, R_N \rangle$. Let C be a minimal conflicting set in R. Let E be a consistent extension of $(D_R, W \cup R_N^*)$. Then,*

1. *if $D_{max(C)} \cup D_{inf(C) \cap R_D} \subseteq GD(E, D)$ then $D_{min(C)} \nsubseteq GD(E, D)$,*

2. *if $D_{min(C)} \cup D_{inf(C) \cap R_D} \subseteq GD(E, D)$ then $D_{max(C)} \nsubseteq GD(E, D)$.*

As with Theorems 3, 4, and 5, the last result refers to the more abstract conception of minimal conflicting sets, as described in Definition 3. The above theorem then can be seen as naturally extending these results to default logic. Observe that only

the defeasible rules in $inf(C)$ are transformed into default rules; the strict rules in $inf(C)$ are dealt with via R_N^*.

Let us relate this theorem to the underlying idea of specificity. Observe that in the first case, where $D_{max(C)} \cup D_{inf(C) \cap R_D} \subseteq GD(E, D)$, 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 $D_{min(C)} \cup D_{inf(C) \cap R_D} \subseteq GD(E, D)$, the less specific defaults apply only if the more specific defaults do not contribute to the given extension.

As regards the specific type of minimal conflicting sets induced by the notion of a conflicting core, we obtain the following result.

Theorem 9 *Let $\langle \langle R_D, R_N \rangle, W \rangle$ be a world description with $R = \langle R_D, R_N \rangle$. Let $(min(C), max(C))$ be the conflicting core of some minimal conflicting set C in R. Let E be a consistent extension of $(D_R, W \cup R_N^*)$. 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 $D_{inf(C) \cap R_D}$ and $inf(C) \cap R_N$.

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 default logic.

Theorem 10 *Let $\langle \langle R_D, R_N \rangle, W \rangle$ be a world description and let E be a set of formulas. Let*

$$D_R^n = \left\{ \frac{\alpha_r : \beta_r}{\beta_r} \;\middle|\; \alpha_r \to \beta_r \in R_D \right\}$$

(and R_r and D_R as in Definition 6).

Define $E_0 = W \cup R_N^*$ and for $i \geq 0$

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

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

It is interesting to note that in determining minimal conflicting sets and conflicting cores, the only place where we distinguish elements of R_N from R_D is in the formation of a conflicting core. We claimed in Section 3.3 that, for the definition of a conflicting core, the reference to R_D is present only to eliminate redundant terms; otherwise we obtain the same default theory. Now that we have the final translation into default logic we can formalise this claim: A *mixed conflicting core* has the same definition as a conflicting core, except that we replace conditions 1. and 2. in Definition 4 by

1. $min(C) \subseteq C_0$,

2. $max(C) \subseteq C_1$.

Then, we have the following result.

Theorem 11 *Let $\langle\langle R_D, R_N\rangle, W\rangle$ be a world description with $R = \langle R_D, R_N\rangle$. Then E is an extension of the corresponding **Z**-default theory iff E is an extension of the default theory obtained in transformation (15) but making appeal to mixed conflicting cores.*

4.3 Discussion of Z-Default Theories

We illustrate the approach further first with an example from [Fahlman *et al.*, 1981], studied in detail by Etherington and Reiter in [1983], and second, with an example involving two minimal conflicting rules. First, we have the following rules (represented by the figure on the right):

Molluscs are normally shell-bearers. $\qquad\qquad M \to S$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Uparrow$
Cephalopods must be Molluscs but normally are *not* shell-bearers. $\qquad C \to \neg S$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Uparrow$
Nautili must be Cephalopods and must be shell-bearers. $\qquad\qquad N \Rightarrow S$

This results in the following generic world description:

$$\begin{aligned} R_D &= \{M \to S,\ C \to \neg S\} \\ R_N &= \{C \Rightarrow M,\ N \Rightarrow C,\ N \Rightarrow S\} \end{aligned}$$

We obtain a single minimal conflicting set C, here expressed as a **Z**-order:

$$\begin{aligned} C_0 &= \{M \to S\} \\ C_1 &= \{C \to \neg S,\ C \Rightarrow M\} \end{aligned}$$

25

The minimal conflicting set contains the conflicting core:

$$(\{M \rightarrow S\}, \{C \rightarrow \neg S\}).$$

As a result, we get: $R_{M \rightarrow S} = \{C \rightarrow \neg S\}$.

Now, given the contingent fact N, we obtain the following default theory:

$$\left(\left\{ \frac{M : S \wedge (C \supset \neg S)}{S}, \frac{C : \neg S}{\neg S} \right\}, \{N\} \cup \{C \supset M, \ N \supset C, \ N \supset S\} \right) \qquad (16)$$

The semi-normal default can be simplified to: $\frac{M : S \wedge \neg C}{S}$. The default theory in (16) has a unique extension in which a Nautilus N is also a Cephalopod C, a Mollusc M, and a shell-bearer S.

Interestingly, default theory (16) is not equivalent to that obtained in [Etherington and Reiter, 1983]. Where they have the semi-normal default $\frac{C : \neg S \wedge (N \supset S)}{\neg S}$ we have the normal default $\frac{C : \neg S}{\neg S}$. However, due to the necessary knowledge $N \supset S$, which is present in all initial sets of facts W, these defaults are equivalent in that one is applicable whenever the other is. Hence our approach avoids the introduction of a redundant semi-normal default. Of course if we replaced the necessary implication $N \Rightarrow S$ with its default counterpart $N \rightarrow S$ we would obtain a second minimal conflicting set and a second conflicting core, and so in this case obtain the semi-normal default. Observe too that there is a second minimal conflicting set, given by the rules $C \rightarrow \neg S, \ N \Rightarrow C, \ \text{and} \ N \Rightarrow S,$ which is ruled out, however, due to its lack of a conflicting core.

Etherington and Reiter start from a network representation comprising "hard", "default" and "exception links". That is, while their network contains "default links" corresponding to $M \rightarrow S$ and $C \rightarrow \neg S$, both require so-called "exception links" indicating that Cephalopods, C, are exceptions to the first default while Nautili, N, are exceptions to the second. These "default links" along with their "exception links" are translated into semi-normal default rules. In this way, the exceptional cases are encoded in the network representation "by hand" in advance. In contrast, we start from a rule-based representation distinguishing strict and defeasible rules; no exceptions are specified. Conflicting default rules are automatically detected by the techniques developed in Section 3 and then mapped onto a semi-normal default theory.

Consider next example (8/9), where we have a minimal conflicting set with more than one minimal conflicting rule.

$$R_0 \ = \ \{A \rightarrow \neg B, C \rightarrow \neg D\} \qquad (17)$$
$$R_1 \ = \ \{A \wedge C \rightarrow B \vee D\} \qquad (18)$$

If we were to represent this as a normal default theory, as described in Section 4.1, then with $W = \{A, C\}$ 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.

The rules in $R_0 \cup R_1$ form a minimal conflicting set with two minimal conflicting rules. This minimal conflicting set comprises two less specific conflicting rules, a situation frequently encountered in multiple inheritance networks. Our approach yields two semi-normal defaults

$$\begin{array}{cc} \dfrac{A\,:\,\neg B \wedge (A \wedge C \supset B \vee D)}{\neg B} \\[6pt] \dfrac{C\,:\,\neg D \wedge (A \wedge C \supset B \vee D)}{\neg D} \end{array} \quad \begin{array}{c} \text{or} \\[6pt] \text{or} \end{array} \quad \begin{array}{c} \dfrac{A\,:\,\neg B \wedge (C \supset D)}{\neg B} \\[6pt] \dfrac{C\,:\,\neg D \wedge (A \supset B)}{\neg D} \end{array} \quad \text{and}$$

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

$$\dfrac{A\,:\,\neg B \wedge (\neg A \vee \neg C)}{\neg B} \quad \text{or} \quad \dfrac{A\,:\,\neg B \wedge (A \wedge C \supset B \vee D) \wedge (C \supset \neg D)}{\neg B} \;,$$

both of which simplify 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

$$\dfrac{A\,:\,\neg B \wedge (C \supset D)}{\neg B} \;,$$

which blocks the second semi-normal default rule in a more subtle way, and additionally allows us to conclude $\neg B$ from $\{A, C, D\}$.

We now examine the formal properties of **Z**-default theories. In regular default logic, 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 whereas semi-normal default theories do not. Transposed to our case, semi-monotonicity 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 default $\frac{P\,:\,\neg F}{\neg F}$ and replace 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]

The existence of extensions is not guaranteed for **Z**-default theories. Consider the rules:

$$A \wedge Q \;\rightarrow\; \neg P, \quad B \wedge R \;\rightarrow\; \neg Q, \quad C \wedge P \;\rightarrow\; \neg R,$$
$$A \qquad \rightarrow \quad P, \quad B \qquad \rightarrow \quad Q, \quad C \qquad \rightarrow \quad R.$$

Each column gives a minimal conflicting set in which the upper rule is more specific than the lower rule. We obtain the rules

$$\frac{A \wedge Q : \neg P}{\neg P}, \qquad\qquad \frac{B \wedge R : \neg Q}{\neg Q}, \qquad\qquad \frac{C \wedge P : \neg R}{\neg R},$$
$$\frac{A : P \wedge \neg Q}{P}, \qquad\qquad \frac{B : Q \wedge \neg R}{Q}, \qquad\qquad \frac{C : R \wedge \neg P}{R}.$$

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 4.5.2.

Another important property is cumulativity. The intuitive idea is that if a theorem is added to the premises from which it was derived, then the set of derivable formulas should remain unchanged. This property is only enjoyed by prerequisite-free normal default theories in regular default logic. It does not hold for **Z**-default theories, as the next example illustrates. Consider the rules $\{D \rightarrow A, A \rightarrow B, B \rightarrow \neg A\}$. The last two rules form a minimal conflicting set. Transforming these rules into defaults, yields $\frac{D : A}{A}, \frac{A : B}{B}, \frac{B : \neg A \wedge (A \supset B)}{\neg A}$, or in the last case $\frac{B : \neg A}{\neg A}$. 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 in passing 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.)

Lastly we show that the translation to obtain **Z**-default theories does not simply reduce the number of extensions obtained in the corresponding normal default theory but may also provide different conclusions. Consider the following world description, where $P \rightarrow S$ stands for "penguins swim".

$$\Delta = \langle\langle\{B \rightarrow F, \ B \rightarrow W, \ P \rightarrow S\}, \{P \Rightarrow B, \ F \Rightarrow \neg S\}\rangle, \{P \wedge \neg S\}\rangle.$$

---

[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'.

While naïve transformation (14) yields normal default theory

$$\left(\left\{ \frac{B:F}{F}, \ \frac{B:W}{W}, \ \frac{P:S}{S} \right\}, \{P \supset B, \ F \supset \neg S\} \cup \{P \wedge \neg S\} \right),$$

transformation (15) results in **Z**-default theory:

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

From the naïve theory, we get an extension containing P, B, W and $\neg S, F$. In contrast, our translation yields an extension including P, B, W and $\neg S$ only; no mention is made of F. This shows that **Z**-theories do not simply eliminate extensions obtained through the naïve transformation; they may even supply us with different conclusions. Intuitively, our **Z**-theories have a conservative attitude towards inheritance over conflicting properties. While there is inheritance of the uncontroversial property W, this is not the case for the controversial property F, which is usually not enjoyed by penguins regardless of swimming ability.

4.4 An Alternative Translation into Default Logic

In Reiter's default logic, a default rule $\alpha \rightarrow \beta$ is informally interpreted as "if α then by default β". However we can also interpret a rule as "by default, if α then β". In this case it is the conditional that is concluded by default, and not the consequent in the presence of the antecedent. In this second interpretation, we turn rules like $\alpha \rightarrow \beta$ into prerequisite-free default rules: For translating rules along with their specificity into prerequisite-free default theories, we replace the definition of δ_r in Definition 6 by

$$\zeta_r = \frac{: (\alpha_r \supset \beta_r) \wedge \bigwedge_{r' \in R_r} (\alpha_{r'} \supset \beta_{r'})}{(\alpha_r \supset \beta_r)} \tag{19}$$

With this transformation, our birds example in (1) together with the knowledge that P is true yields default theory:

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

From this theory, we obtain a single extension containing $\{P, \neg F, B, W\}$.

As discussed in [Delgrande *et al.*, 1994], the problem of controlling interactions among such prerequisite-free default rules is more acute than in the regular case. Consider our initial example (1) and turn the implication $P \Rightarrow B$ into its default counterpart $P \rightarrow B$. The usual translation, ignoring specificity information, translates this into the following prerequisite-free default rules:

$$\frac{:B \supset F}{B \supset F}, \ \frac{:B \supset W}{B \supset W}, \ \frac{:P \supset B}{P \supset B}, \ \frac{:P \supset \neg F}{P \supset \neg F} \tag{20}$$

Given P, we obtain three extensions, containing $\{P, \neg F, B, W\}$, $\{P, F, B, W\}$, and $\{P, \neg F, \neg B\}$.[9] In the regular default theory (with prerequisites) we obtain just the first two extensions. Clearly, transformation (19) eliminates the second extension. The third extension however remains; moreover this extension hinders property inheritance, since we cannot conclude that birds have wings. This is caused by the contrapositive of $B \to F$. 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: by strengthening the blocking conditions for minimal conflicting rules or by blocking the contrapositive of minimal conflicting rules. In the first case, we could turn $B \to F$ into $\frac{: (B \supset F) \wedge \neg P}{B \supset F}$ by adding the negated antecedents of the maximal conflicting rules, here $\neg P$. While this looks appealing, we have already seen in Section 4.3 that this is too strong in the presence of multiple minimal conflicting rules. To see this, consider the rules given in (8/9). For $A \to \neg B$, we obtain $\frac{: (A \supset \neg B) \wedge (\neg A \vee \neg C)}{A \supset \neg B}$ or $\frac{: A \supset (\neg B \wedge \neg C)}{A \supset \neg B}$. As argued in Section 4.3, there is no reason why $A \to \neg B$ should not be applied given the facts $\{A, C, D\}$. Also, in general it does not make sense to address a problem stemming from contrapositives 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 $B \to F$ 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 $B \to F$ 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 $A \to \neg B$ in (8/9), 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 can alternatively replace the definition of δ_r in Definition 6 by[10]

$$\zeta_r' = \frac{: (\alpha_r \supset \beta_r) \wedge \beta_r \wedge \bigwedge_{r' \in R_r} (\alpha_{r'} \supset \beta_{r'})}{(\alpha_r \supset \beta_r)}. \tag{21}$$

Applying this transformation to the set of rules in (20), we obtain:

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

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

---

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

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

Note that blocking the contrapositive of minimal conflicting rules is an option outside the presented framework. The purpose of the above transformation is to preserve inheritance over default statements such as $P \to B$. Inheritance over strict statements, like $P \Rightarrow B$, however can be done *without* blocking contrapositives. In this case, of course, transformation (19) is sufficient.

Transformations (19/21) 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 classical conditionals unless explicitly blocked. However, the counterexamples for semi-monotonicity, cumulativity, and the existence of extensions carry over to prerequisite-free **Z**-default theories.

4.5 Further Translations and Comparison with Related Approaches

In a manner similar to the approach described in the previous sections, we can compile prioritised rules into variants of default logic, including those of [Baader and Hollunder, 1993a, Brewka, 1993, Boutilier, 1992b], as well as Theorist [Poole, 1988] and autoepistemic logic [Moore, 1985]. This is described in the remainder of this section.

4.5.1 Ordered Variants of Default Logic

At the start of this section we described how to extract a strict partial order from a family of minimal conflicting sets 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 a general interface between approaches, in that we can also use 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 6. We do this by associating with each normal default $\frac{\alpha \,:\, \beta}{\beta}$ a rule $\alpha \to \beta$ 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 (15) to turn ordered normal default theories into semi-normal 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 default logic is modified. In brief, a default is only applicable at an iteration step (in the sense of Definition 5) if no more specific (or $<$-greater) default is applicable.[11] 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

---

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

a total extension of $<$, Baader and Hollunder allow for applying all $<$-maximal defaults at each step.

As a first example, consider the normal default rules

$$\frac{:A}{A}, \quad \frac{:B}{B}, \quad \frac{B:C}{C}, \quad \frac{A:\neg C}{\neg C}$$

(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 in their approach 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

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

In regular default logic, the resultant default theory yields only the first extension containing $\{A, B, C\}$. So here our approach yields the same result as Baader and Hollunder's approach.

As a second example, again from [Baader and Hollunder, 1993b], consider the rules

$$\frac{:A}{A}, \quad \frac{B:\neg A}{\neg A}, \quad \frac{:B}{B}, \quad \frac{A:\neg B}{\neg B}$$

(for short $\delta_1, \delta_2, \delta_3, \delta_4$), along with $\delta_1 < \delta_2$, $\delta_3 < \delta_4$. Baader and Hollunder 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

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

which yields only the first two extensions in default logic. Thus, as opposed to the previous example, our approach yields here the same result as Brewka's approach. 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.

Our approach to compiling partial orders into semi-normal default theories assumes that priorities are determined by specificity conflicts between rules. Consider where we might extract priorities directly from subsumption relations, as is done in [Baader and Hollunder, 1993a]. Consider terms stating that "birds fly", $B \rightarrow F$, and "young birds need special care", $B \wedge Y \rightarrow C$, along with the subsumption relation between "birds" and "young birds". This subsumption amounts

to a priority between the two rules even though there is no conflict: $(B \rightarrow F) < (B \wedge Y \rightarrow C)$. This priority would not be detected in our approach, since we rely on there being a conflict between rules to determine priorities. In this example there is no minimal conflicting set and so in our approach we would obtain the two normal rules $\frac{B\,:\,F}{F}$ and $\frac{B \wedge Y\,:\,C}{C}$.

The preceding exposition assumed that a rule $\alpha \rightarrow \beta$ was associated with a default having prerequisite α and consequent β. This view underlies [Baader and Hollunder, 1993a] and [Brewka, 1993], in that 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 showed in the last subsection. [Boutilier, 1992b] proposes an approach based on these intuitions, where a ranking on defaults is obtained from the **Z**-ordering of the defaults. Boutilier 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 default logic by applying the rules in the highest set, and working down. However this means that for reasoning, one again steps outside the machinery of default logic. Moreover, since the order in which defaults are applied depends on the original **Z**-ordering, this order may be "upset" by the addition of irrelevant conditionals. This in turn may introduce unwanted priorities. For this last point, consider again example (8/9), but where we also have that A is exceptional with respect to E for some property A'. We had originally

$$
\begin{aligned}
(R_0, R_1) \quad = \quad & (\{A \rightarrow \neg B, C \rightarrow \neg D\}, \{A \wedge C \rightarrow B \vee D\}) \quad \text{and to this we add} \\
& \{A \rightarrow \neg E,\ A \rightarrow A',\ A' \rightarrow E\}.
\end{aligned}
$$

This yields the **Z**-ordering, expressed as a union of sets to clarify the structure:

$$
\begin{aligned}
R_0 \quad &= \quad \{C \rightarrow \neg D\} \cup \{A' \rightarrow E\} \\
R_1 \quad &= \quad \{A \wedge C \rightarrow B \vee D,\ A \rightarrow \neg B\} \cup \{A \rightarrow \neg E,\ A \rightarrow A'\}
\end{aligned}
$$

Intuitively A' and E have nothing to do with the original theory, yet their addition has "moved" $A \rightarrow \neg B$ from R_0 to R_1. In Boutilier's approach we would conclude $\neg B$ and D from $\{A, C\}$, and would not get the second extension as in the original case. Hence adding irrelevant defaults leads to different extensions wrt applying the original three defaults. This phenomenon clearly is avoided in the present approach, due to our use of minimal conflicting set and not the full **Z**-ordering.

4.5.2 Other Variants of Default Logic

Another alternative is the translation into variants of default logic that guarantee the existence of extensions [Łukaszewicz, 1988, Brewka, 1991, Delgrande *et al.*, 1994]. This can be accomplished by means of both translation (15) and (19/21). Moreover, the resulting **Z**-default theories enjoy cumulativity when applying translation (15) and (19/21) in the case of cumulative default logic and when applying translation (19/21) 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, they all 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 see [Brewka, 1991].

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 prioritised rules by first applying transformation (19/21) and then that given in [Delgrande *et al.*, 1994] for translating prerequisite-free default theories in constrained default logic into Theorist.

4.5.3 Autoepistemic Logic

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 where a formula $L\alpha$ is 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 "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 $R_r \subseteq R$ (as given in Definition 6); for each $r \in R$ we define:

$$\rho_r \;=\; \alpha_r \wedge \neg L\neg \big(\beta_r \wedge \bigwedge_{r'\in R_r}(\alpha_{r'} \supset \beta_{r'})\big) \supset \beta_r,$$
$$\varrho_r \;=\; L\alpha_r \wedge \neg L\neg \big(\beta_r \wedge \bigwedge_{r'\in R_r}(\alpha_{r'} \supset \beta_{r'})\big) \supset \beta_r.$$

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

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

along with $B \land \neg L \neg W \supset W, P \land \neg L \neg B \supset B$, and $P \land \neg LF \supset \neg F$ 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.

5 Compiling Specificity into Minimization-Based Approaches

5.1 Circumscription

Circumscription was introduced by John McCarthy in [1980, 1986] as an approach to formalizing diverse nonmonotonic aspects of commonsense reasoning. The idea behind circumscription is that of "logical minimization". A formula α follows from a theory W by circumscription if α is true in all models of W that are minimal in a certain sense. In applications to default reasoning, circumscription is used to minimise "abnormalities" of default rules. For this purpose, the language is enriched by abnormality propositions designated $ab_1, ab_2, \ldots$ These propositions address cases exceptional to a default rule at hand. In this way, a default statement like "birds fly" is represented as a classical implication of the form $B \land \neg ab \supset F$. Following this general principle, we transform a set of default rules into a set of implications. For a generic world description $\langle R_D, R_N \rangle$, we define

$$
\begin{align}
R_D^{ab} &= \{\alpha_r \land \neg ab_r \supset \beta_r \mid \alpha_r \rightarrow \beta_r \in R_D\}; \tag{22} \\
R_N^* &= \{\alpha_r \supset \beta_r \mid \alpha_r \Rightarrow \beta_r \in R_N\}. \tag{23}
\end{align}
$$

As an example, consider the following simplification of our initial example in (2):

$$
\langle R_D, R_N \rangle = \langle \{B \rightarrow F,\ P \rightarrow \neg F\}, \{P \Rightarrow B\} \rangle. \tag{24}
$$

For R_D, we obtain:

$$
R_D^{ab} = \{B \land \neg ab_1 \supset F,\ P \land \neg ab_2 \supset \neg F\}.
$$

The strict rule in R_N becomes $R_N^* = \{P \supset B\}$. Now, the set of rules in $R_D^{ab} \cup R_N^*$ can be seen as a description of our birds scenario in (24) in standard propositional logic. However, we also want to express that things are considered as normal as possible—provided that there is no evidence to the contrary. This assumption is formally accomplished by circumscribing a world description: Let W be a propositional formula, and $P \dot\cup Z$ a partition of all atoms in W. The circumscription of P in W while varying Z is defined as (cf. [McCarthy, 1980, Lifschitz, 1985])

$$
Circum(W; P; Z) = W \land \left(\forall P', Z'\ (W[P/P', Z/Z'] \land (P' \supset P) \supset (P \supset P')) \right).
$$

In this formula, P' and Z' are disjoint sets of new propositional variables corresponding to those in P and Z. That is, $P' = \{p' \mid p \in P\}$ and $Z' = \{z' \mid z \in Z\}$. The formula $W[P/P', Z/Z']$ denotes the result obtained by replacing in W all occurrences of variables in $P \cup Z$ by their counterparts in $P' \cup Z'$. $(P' \supset P)$ and $(P \supset P')$ abbreviate $\bigwedge_{p \in P}(p' \supset p)$ and $\bigwedge_{p \in P}(p \supset p')$, respectively. The net result is that the circumscription axiom asserts that the number of atoms in P that are true is as small as possible. Furthermore, in achieving this minimisation, the truth values of atoms in Z are allowed to vary. The semantical underpinnings for circumscription are given by minimal models. For interpretations[12] M and N, we define $M \leq_{(P;Z)} N$ if $M \cap P \subseteq N \cap P$. Then, according to [Lifschitz, 1985], M is a model of *Circum*$(W; P; Z)$ iff M is minimal among all models of W with respect to $\leq_{(P;Z)}$.

Consider our birds example in (24) along with fact P. The models of $\{P\}$, R_N^*, and R_D^{ab}, where we just list the positive literals, are the following:

$$\{P, B, F, ab_1, ab_2\}, \; \{P, B, ab_1, ab_2\}, \; \{P, B, F, ab_2\}, \; \{P, B, ab_1\} \qquad (25)$$

Circumscribing propositions ab_1 and ab_2 while varying B, F, P amounts to reasoning with respect to those models in (25) that have the fewest abnormality propositions. There are two such minimal models, $\{P, B, ab_1\}$ and $\{P, B, F, ab_2\}$. As a consequence, we cannot conclude much more than from the original world description in (24). That is, we have

$$\textit{Circum}(\{P\} \cup R_N^* \cup R_D^{ab}\,;\, \{ab_1, ab_2\}\,;\, \{B, F, P\}) \models B \wedge P \wedge (ab_1 \equiv \neg ab_2).$$

In particular, we cannot derive ab_1, ab_2, or F, nor their negation. This shows that circumscription does not respect the principle of specificity. This shortcoming was observed in [McCarthy, 1986]. For fixing this problem, McCarthy introduced a prioritised version of circumscription by assuming that abnormality propositions are a priori assigned different priorities. However, in this approach specificity is handled on the metalevel by iterating circumscription on certain priority layers (and so by extrasystematic means). We examine this approach in Section 5.3.

5.2   Z-Circumscription Theories

We address the lack of specificity by providing axioms reflecting the precedence of more specific rules over less specific rules. We accomplish this, again, by taking advantage of the specificity information provided by minimal conflicting sets. This information is encoded by means of axioms that express precedences among default rules conflicting because of differing specificity.[13]

---

[12]We represent interpretations as sets of atoms.
[13]Compare Definition 6.

Definition 7 *Let* $R = \langle R_D, R_N \rangle$ *be a generic world description. Let* $(C^i)_{i \in I}$ *be the family of all minimal conflicting sets in R. We define*

$$R_D^{sp} = \{\neg\bigwedge_{r' \in R_r}(\alpha_{r'} \supset \beta_{r'}) \supset ab_r \mid r \in R_D\} \tag{26}$$

where $R_r = \{r' \in max(C^i) \mid r \in min(C^i) \, for \, i \in I \, \}$.

This definition provides a *circumscription policy* for specificity using standard circumscription, rather than prioritised circumscription [McCarthy, 1986]. Observe that rules like $\neg\bigwedge_{r' \in R_r}(\alpha_{r'} \supset \beta_{r'}) \supset ab_r$ have the form of *inheritance cancellation axioms* as described in [McCarthy, 1986]. Note that necessary rules are discarded in the formation of antecedents of specificity axioms, since they constitute true sentences in the world description obtained by adding R_N^*. Consequently, a specificity axiom is tautological if R_r consists of necessary rules only. We detail the effect of specificity axioms below.

Consider our birds example in (24). As required in Definition 7, we associate with each default rule a set of more specific, conflicting default rules extracted from the minimal conflicting sets. For the default rules R_D in (24), this yields the following sets of default rules according to the specification of R_r in Definition 7:

$$R_{B \to F} = \{P \to \neg F\} \qquad\qquad R_{P \to \neg F} = \emptyset$$

Recall that the first set expresses the fact that $P \to \neg F$ is more specific than $B \to F$, while the second equation tells us that there is no more specific (conflicting) default rule than $P \to \neg F$. According to Definition 7, we obtain for the default rules R_D in (24) the following specificity axiom:

$$R_D^{sp} = \{\neg(P \supset \neg F) \supset ab_1\}. \tag{27}$$

We have omitted the tautology obtained in the case of $r = P \to \neg F$.

Finally, our construction yields the following classical theory in propositional logic when applied to our generic world description in (24):

$$R_D^{ab} \cup R_D^{sp} \cup R_N^* = \{B \wedge \neg ab_1 \supset F, \, P \wedge \neg ab_2 \supset \neg F\} \cup \{\neg(P \supset \neg F) \supset ab_1\} \cup \{P \supset B\}. \tag{28}$$

Together with the contingent fact P, these rules have the following models:

$$\{P, B, F, ab_1, ab_2\}, \, \{P, B, ab_1, ab_2\}, \, \{P, B, ab_1\} \tag{29}$$

In contrast to the models in (25), we now have only a single minimal model, namely $\{P, B, ab_1\}$. Hence, circumscription allows for concluding P, B, ab_1, and $\neg F$, $\neg ab_2$ from the transformed world description. That is,

$$Circum(\{P\} \cup R_D^{ab} \cup R_D^{sp} \cup R_N^* \, ; \, \{ab_1, ab_2\} \, ; \, \{B, F, P\}) \models P \wedge B \wedge \neg F.$$

One might wonder why we compose specificity axioms by taking entire rules and not merely their antecedents. This would amount to considering the rule $P \supset ab_1$ instead of the one given in (27). Interestingly, $P \supset ab_1$ is the inheritance cancellation axiom suggested in [McCarthy, 1986]. Let us illustrate this by regarding again theory (8/9), constituting a minimal conflicting set with two less specific conflicting rules:

$$C_0 = \{A \to \neg B, C \to \neg D\} \qquad\qquad C_1 = \{A \wedge C \to B \vee D\}.$$

Applying the transformations in (22) and (26) yields:

$$
\begin{aligned}
R_D^{ab} &= \{(A \wedge \neg ab_1) \supset \neg B, \ (C \wedge \neg ab_2) \supset \neg D, \ (A \wedge C \wedge \neg ab_3) \supset B \vee D\} \\
R_D^{sp} &= \{\neg(A \wedge C \supset B \vee D) \supset ab_1, \ \neg(A \wedge C \supset B \vee D) \supset ab_2\}. \qquad (30)
\end{aligned}
$$

Given $\{A, C, D\}$, we observe a single "abnormality minimal" model of the resulting world description, namely $\{A, C, D, ab_2\}$. In this way, circumscription allows us to conclude $\neg B$. This amounts to applying the first default $A \to \neg B$. Now, let us replace the specificity axioms in (30) by

$$R_D^{ic} = \{A \wedge C \supset ab_1, \ A \wedge C \supset ab_2\}$$

according to the putative recipe described above. We obtain for the world description built on the facts $\{A, C, D\}$ *two* "abnormality minimal" models, $\{A, C, D, ab_1, ab_2\}$ and $\{A, B, C, D, ab_1, ab_2\}$. Consequently, we cannot derive $\neg B$. Given $\{A, C, D\}$ there is, however, no reason why the default $A \to \neg B$ should not apply. This shows that our approach is advantageous over plain blocking conditions. Another advantage of our construction is that it allows for an elegant alternative formulation for incorporating specificity, as we will see next.

Observe that the two implications involving P and $\neg F$ in (28) can be put together to:

$$P \wedge (\neg ab_2 \vee \neg ab_1) \supset \neg F . \qquad (31)$$

This indicates that we can alternatively modify the more specific rules instead of adding axioms referring to the least specific ones. As a general result, the next definition provides us with an alternative but more compact translation of default rules, along with their specificity information, into classical logic:

Definition 8 *Let* $R = \langle R_D, R_N \rangle$ *be a generic world description. Let* $(C^i)_{i \in I}$ *be the family of all minimal conflicting sets in R. We define*

$$R_D^\star = \{\alpha_r \wedge \left(\neg ab_r \vee \bigvee_{r' \in R_r'} \neg ab_{r'}\right) \supset \beta_r \mid r \in R_D\}$$

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

This transformation unifies those given in (22) and (26). This is accomplished by slightly extending transformation (22) in the case of more specific conflicting default rules. To this end, we extend the abnormality condition of more specific default rules by disjoining the abnormality propositions of the less specific default rules. This yields for our generic world description in (24) the following set of strict rules:

$$R_D^\star \cup R_N^* = \{B \wedge \neg ab_1 \supset F,\ P \wedge (\neg ab_2 \vee \neg ab_1) \supset \neg F\} \cup \{P \supset B\}.$$

This result should be compared with that obtained in (28). Note furthermore that R_r' reverses the roles of $min(C^i)$ and $max(C^i)$ in the specification of R_r in Definition 7.

The next theorem tells us that the constructions are in fact equivalent:

Theorem 12 *Let $\langle R_D, R_N \rangle$ be a generic world description. Then, $R_D^{ab} \cup R_D^{sp} \cup R_N^*$ is logically equivalent to $R_D^\star \cup R_N^*$.*

We further study the effect of our approach by means of a general but simple example. Consider a generic world description comprising two conflicting rules r and r' and suppose that r is more specific than r'. Hence, we have $R_r' = \{r'\}$ and we obtain in turn

$$\alpha_r \wedge (\neg ab_r \vee \neg ab_{r'}) \supset \beta_r \qquad \text{and} \qquad \alpha_{r'} \wedge \neg ab_{r'} \supset \beta_{r'}\ .$$

Our intended interpretation is that r is preferable over r' (because of specificity) whenever r and r' are both potentially "applicable". Assume that β_r and $\beta_{r'}$ are not jointly satisfiable, since r and r' conflict. Then, the first default takes precedence over the second whenever both antecedents, α_r and $\alpha_{r'}$, are derivable. This is so because of the following reasons: Clearly, $\{\alpha_r, \alpha_{r'}, \beta_r, \beta_{r'}\}$ is no "abnormality minimal" model, since β_r and $\beta_{r'}$ are not jointly satisfiable. Also,

$$\{\alpha_r, \alpha_{r'}, ab_r, \beta_{r'}\}, \{\alpha_r, \alpha_{r'}, ab_r, \beta_r\}, \{\alpha_r, \alpha_{r'}, ab_r\}, \{\alpha_r, \alpha_{r'}, ab_{r'}, \beta_{r'}\}, \{\alpha_r, \alpha_{r'}, ab_{r'}\}$$

are not "abnormality minimal" models, since they either falsify $\alpha_r \wedge (\neg ab_r \vee \neg ab_{r'}) \supset \beta_r$ or $\alpha_{r'} \wedge \neg ab_{r'} \supset \beta_{r'}$. Finally, there remain three candidates satisfying a single abnormality proposition:

$$\{\alpha_r, \alpha_{r'}, ab_r, \beta_r, \beta_{r'}\},\ \{\alpha_r, \alpha_{r'}, ab_{r'}, \beta_r, \beta_{r'}\},\ \{\alpha_r, \alpha_{r'}, ab_{r'}, \beta_r\}.$$

In fact, these three models are "abnormality minimal". Since all of them satisfy β_r, they prefer the more specific default r over r'.

In analogy to Theorems 8 and 9, we have the following general results. Theorem 13 is with respect to the general theory of minimal conflicting sets while Theorem 14 is with respect to the specific development involving conflicting cores.

Theorem 13 *Let $\langle \langle R_D, R_N \rangle, W \rangle$ be a world description with $R = \langle R_D, R_N \rangle$. Let C be a minimal conflicting set in R. Then,*

1. *if $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \models inf(C)^* \wedge \left(\bigwedge_{r \in max(C)} \alpha_r \wedge \neg ab_r \right)$ then $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \not\models \left(\bigwedge_{r \in min(C)} \beta_r \right)$*

2. *if $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \models inf(C)^* \wedge \left(\bigwedge_{r \in min(C)} \beta_r \right)$ then $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \not\models \left(\bigwedge_{r \in max(C)} \alpha_r \wedge \neg ab_r \right)$*

where Z is the set of all propositional variables in W, R_D, R_N.

In the first case, we have that

$$Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \models \left(\bigwedge_{r \in min(C)} \alpha_r \right)$$

by Theorem 3. So though the prerequisites of the minimal conflicting defaults are derivable, the corresponding rules do not apply. This is so because some of the abnormality propositions ab corresponding to the minimal conflicting defaults must hold. In this way, the more specific defaults from $max(C)$ take precedence over the less specific defaults in $min(C)$. Conversely, in the second case, the less specific defaults apply only if the more specific defaults do not contribute to the given extension.

We also obtain the following result.

Theorem 14 *Let $\langle \langle R_D, R_N \rangle, W \rangle$ be a world description with $R = \langle R_D, R_N \rangle$. Let $(min(C), max(C))$ be the conflicting core of some minimal conflicting set C in R. Then,*

1. *if $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \models \left(\bigwedge_{r \in max(C)} \alpha_r \wedge \neg ab_r \right)$ then $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \not\models \left(\bigwedge_{r \in min(C)} \alpha_r \wedge \beta_r \right)$*

2. *if $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \models \left(\bigwedge_{r \in min(C)} \alpha_r \wedge \beta_r \right)$ then $Circum(W \cup R_D^\star \cup R_N^* \,;\, \{ab_r \mid r \in R_D\} \,;\, Z) \not\models \left(\bigwedge_{r \in max(C)} \alpha_r \wedge \neg ab_r \right)$*

where Z is the set of all propositional variables in W, R_D, R_N.

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

Importantly, both transformations are consistency-preserving.

Theorem 15 *Let $\langle R_D, R_N \rangle$ be a generic world description such that $R_N^* \cup R_D^*$ is satisfiable. Then, $R_D^\star \cup R_N^*$ is satisfiable.*

Clearly, the same result applies to our initial approach using the rules in $R_D^{ab} \cup R_D^{sp} \cup R_N^*$. Moreover, consistency is also preserved when applying circumscription to specificity-integrating world descriptions:

Theorem 16 *Let $\langle\langle R_D, R_N \rangle, W \rangle$ be a world description such that $W \cup R_N^*$ is satisfiable. Then, $Circum(W \cup R_D^\star \cup R_N^* ; \{ab_r \mid r \in R_D\} ; Z)$ is satisfiable, where Z is the set of all propositional variables in W, R_D, R_N.*

Thus, unlike standard default logic (and autoepistemic logic; see Section 4.4) in which one might obtain incoherent theories, the transformation given for circumscription cannot render an original theory incoherent.

Finally, Theorem 16 provides us with a compact summary of our approach. That is, we start from knowledge bases of the form $\langle\langle R_D, R_N \rangle, W \rangle$. Next, we treat the generic part of such a world description by means of techniques developed in Section 3 for isolating minimal sets of default rules with specificity conflicts. Then, we take the initial world description along with the determined specificity information obtained in the previous step and translate it into a classical theory in propositional logic, namely $W \cup R_D^\star \cup R_N^*$. In a final step, we compute the resulting conclusions by circumscribing the introduced abnormality propositions while varying all propositions in the original world description.

5.3 Discussion and Related Work

McCarthy [1986] addressed the lack of specificity in circumscription by introducing *prioritised circumscription*. This approach has been further developed in [Lifschitz, 1985]. The idea is to partition the circumscribed propositions into priority layers. For this, a set of propositions P is partitioned into disjoint subsets $P_1 \dot\cup \ldots \dot\cup P_n$ where the propositions in P_i should take priority over those in P_{i+1}. For computing the result of prioritised circumscription, Lifschitz shows in [1985] that any prioritised circumscription, written $Circum(W; P_1 > \ldots > P_n; Z)$, can be expressed as a conjunction of ordinary circumscriptions. That is, $Circum(W; P_1 > \ldots > P_n; Z)$ is equivalent to $\bigwedge_{i=1}^{n} Circum(W; P_i ; P_{i+1} \cup \ldots \cup P_n \cup Z)$. Thus a prioritised circumscription becomes a sequence of ordinary circumscriptions where propositions in a higher layer are "minimised" while propositions in a lower layer are varied. This process is iterated over all layers in the partition of P.

In our simple birds example, we could give the abnormality proposition of the more specific default rule ab_2 priority over the one of the less specific rule ab_1. This yields

$$Circum(\{P\} \cup R_D^{ab} \cup R_N^* ; \{ab_2\} > \{ab_1\} ; \{B, F, P\}) =$$

$$Circum(\{P\} \cup R_D^{ab} \cup R_N^* \,;\, \{ab_2\};\, \{ab_1, B, F, P\})$$
$$\wedge \quad Circum(\{P\} \cup R_D^{ab} \cup R_N^* \,;\, \{ab_1\}\,;\, \{B, F, P\}) \qquad \models P \wedge B \wedge \neg F.$$

Observe however that such an approach works only if the partial order induced by specificity can be pressed into a "layered format" without introducing unwanted preference relations. For instance, [Grosof, 1991] gives partial orders that cannot be represented in a "layered format". For similar reasons one usually refrains from mapping a full **Z**-ordering onto "layered structures" in default reasoning systems (see Section 4.5). Finally, the iterated format used for computing prioritised circumscription amounts to a treatment of specificity on the metalevel, rather than producing an object level theory, as in our approach.

Grosof [1991] extends prioritised circumscription for dealing with partial orders. In this way, his approach allows for formalizing partial orders representing specificity information. There are however some important differences to our approach. First, [Grosof, 1991] describes default rules with specificity in terms of an extended prioritised circumscription, while we deal with the basic approach to circumscription. Second, it is (to our knowledge) yet unknown how and if Grosof's extended prioritised circumscription is reducible to iterated ordinary circumscriptions. Hence his approach remains outside the basic circumscriptive machinery. Consequently, it is impossible to apply, for instance, techniques for transforming circumscription in first-order or even propositional logic (see [Lifschitz, 1985]). Observe that, in contrast, we deal with a single basic circumscriptive theory that allows for applying the aforementioned techniques along with existing implementations of circumscriptive theorem provers [Ginsberg, 1989].

The choice of circumscription as a target formalism rather than, say, default logic or autoepistemic logic, has several benefits. Foremost, we work largely within classical logic: a classical theory is generated; there is a circumscriptive step minimising ab sentences, after which one has a classical knowledge base. Circumscription itself has several rather nice features: we don't have the notion of "extension" to be concerned with; also circumscription is cumulative and consistency-preserving. Furthermore computational properties of circumscription are well-studied [Lifschitz, 1985, Ginsberg, 1989].

Circumscription and its variants are by far the best known and best studied minimization-based approach to nonmonotonic reasoning. Consequently we have little to say about other approaches. Predicate completion [Clark, 1978] for one has seen little application to areas of reasoning with default properties. For the closed world assumptions, presumably the only appropriate variant of the original approach is the *careful* closed world assumption [Eiter and Gottlob, 1993], which is subsumed by parameterised circumscription. Hence here the compilation is the same as that as given for parameterised circumscription. Shoham's *preferential*

entailment [Shoham, 1988], which is based on minimising certain models, in fact it has been shown to have strong ties with the conditional logic corresponding to S4 and, so in its most general conception, enforces specificity [Lamarre, 1991].

6 Discussion

This paper has addressed the issue of incorporating specificity information into those approaches to nonmonotonic reasoning where specificity conflicts are not handled as part of the basic machinery of the approach; these approaches have generally been characterised as either "consistency-based" or "minimization-based". We begin with a generic world description expressed as a set of strict and defeasible rules. The end result is a default theory expressed in some (consistency-based or minimization-based) formalism, and where conflicts involving differing specificities are resolved. By appeal to a theory of defaults, here using the notion of toleration from System **Z**, specificity conflicts are isolated into minimal conflicting sets. From these minimal conflicting sets the conflicting rules are determined, and from the specificity information intrinsic in these sets, a default theory in a target language in turn is specified. For default logic the end result is a semi-normal default theory; in circumscription the end result is a set of abnormality propositions that, when circumscribed, yield a theory in which specificity information is appropriately handled. While we mainly deal with theories expressed (ultimately) in default logic and circumscription, we also address variants of these approaches as well as autoepistemic logic and Theorist. Arguably the approach is both uniform and general, and so is applicable to any sufficiently rich approach to nonmonotonic reasoning that does not "automatically" deal with specificity conflicts.

This approach is modular, in that we separate the *determination* of conflicts from the *resolution* of conflicts among rules. The approach provides a broadly-applicable framework, subsuming for example that of [Reiter and Criscuolo, 1981]. As well, it generalises related mappings as found for example in [Etherington and Reiter, 1983] and, in circumscription, yields a more comprehensive form of an inheritance cancellation axiom. This work differs from previous work in specifying priorities among default rules (in both default logic and circumscription) in that specificity information is obtained from information intrinsic in the rules, rather than assumed to exist a priori. In contrast to previous work, the approach avoids stepping outside the machinery of the underlying approach. Thus, we deal with the "standard" version of default logic and circumscription, and do not need to rely on prioritised versions, as do other approaches.

With respect to implementation, default reasoning is intractable in the worst case in virtually all conceptions. For example there may be an exponential number

of specificity conflicts in a default theory. Nonetheless our approach would appear to be relatively amenable to implementation. First, the fact that we deal with the standard versions of default logic and circumscription means that one can make use of existing theorem provers in implementing this approach. Second, we have that if a default theory contains a minimal conflicting set, then so does any superset of the theory; this then would allow an incremental computation of minimal conflicting sets, even in evolving knowledge bases.

7   Proofs of theorems

In what follows, we use the following function providing the set of negated elements. Define for a set of formulas S,

$$\overline{S} = \{\neg\alpha \mid \alpha \in S\}.$$

Let us moreover adopt the convention of using interchangeably a finite set of formulas S and the conjunction of its elements $\bigwedge_{\alpha \in S} \alpha$ on both sides of the entailment relation $\models$.

Proof 1 Consider a minimal conflicting set C and assume the converse, that is that C has more than a binary **Z**-ordering. Then, there is a rule $r \in C \setminus (C_0 \cup C_1)$ such that

$$C_1^* \cup \{\alpha_r \wedge \beta_r\}$$

is unsatisfiable. Consider $C_0 \cup C_1$. Obviously, this set has a non-trivial **Z**-ordering. This is a contradiction to the minimality of C. ∎

Proof 2 We have

$$\{\alpha_1 \wedge \beta_1\} \cup C^*$$

is unsatisfiable for all $\alpha_1 \rightarrow \beta_1 \in C_1$.

This implies

$$C^* \models \alpha_1 \supset \neg\beta_1$$

for all $\alpha_1 \rightarrow \beta_1 \in C_1$; since (trivially)

$$C^* \models \alpha_1 \supset \beta_1$$

we obtain by reductio ad absurdum that

$$C^* \models \neg\alpha_1$$

for all $\alpha_1 \rightarrow \beta_1 \in C_1$. ∎

Proof 3 We have

$$inf(C)^* \cup max(C)^* \models Prereq(max(C)) \supset Prereq(min(C))$$

iff

$$inf(C)^* \cup max(C)^* \cup Prereq(max(C)) \cup \overline{Prereq(min(C))}$$

is unsatisfiable.

Let us assume the latter set is satisfiable.

We have (trivially)

$$\overline{Prereq(min(C))} \models min(C)^*.$$

This implies

$$inf(C)^* \cup max(C)^* \cup Prereq(max(C)) \cup \overline{Prereq(min(C))} \models C^*.$$

But according to Theorem 2

$$C^* \models \overline{Prereq(max(C))},$$

a contradiction! ∎

Proof 4 Let us assume the opposite. That is,

$$(R')^* \models Prereq(R'') \supset Prereq(max(C)).$$

Or, equivalently,

$$(R')^* \cup Prereq(R'') \models Prereq(max(C)).$$

According to Theorem 2,

$$C^* \models \overline{Prereq(max(C))}.$$

By monotonicity,

$$(R')^* \cup Prereq(R'') \models \overline{Prereq(max(C))}.$$

Consequently,

$$(R')^* \cup Prereq(R'')$$

is unsatisfiable, a contradiction! ∎

Proof 5 We have

$$inf(C)^* \cup \{\alpha_r\} \models \neg(Conseq(min(C)) \wedge Conseq(max(C)))$$

iff

$$inf(C)^* \cup \{\alpha_r\} \cup Conseq(min(C)) \cup Conseq(max(C))$$

is unsatisfiable.

Let us assume the last set is satisfiable.

Clearly,

$$Conseq(min(C)) \models min(C)^*$$

and

$$Conseq(max(C)) \models max(C)^*.$$

This implies

$$inf(C)^* \cup \{\alpha_r\} \cup Conseq(min(C)) \cup Conseq(max(C)) \models C^*.$$

But according to Theorem 2

$$C^* \models \overline{Prereq(max(C))},$$

a contradiction (since $\alpha_r \in Prereq(max(C))$)! ∎

Proof 6 Given immediately after Theorem 6. ∎

Proof 7 Let $(C_0, C_1) = C$ be a minimal conflicting set.

Since

$$C \cup \{\alpha_r \wedge \beta_r\} \models \bot$$

for $r \in C_1$, it must be that

$$\{\alpha_{r'} \wedge \beta_{r'} \mid r' \in C_0 \cup C_1\} \cup \{\alpha_r \wedge \beta_r\} \models \bot$$

or, since $r \in C_1$, $\{\alpha_{r'} \wedge \beta_{r'} \mid r' \in C_0 \cup C_1\} \models \bot$.

Since by assumption

$$\{\alpha_r \wedge \beta_r \mid r \in C_0\} \not\models \bot,$$

we can essentially begin with $\{\alpha_r \wedge \beta_r \mid r \in C_0\}$ and add rules from C_1 until inconsistency is obtained. ∎

Proof 8 Let R be a set of rules and let W be a set of formulas. Let C be a minimal conflicting set in R. Let E be a consistent extension of (D_R, W).

1. Let $D_{max(C)} \cup D_{inf(C) \cap R_D} \subseteq GD(E, D)$. By definition of $GD(E, D)$, this implies that

$$Prereq\Big(D_{max(C)} \cup D_{inf(C) \cap R_D}\Big) \cup Conseq\Big(D_{max(C)} \cup D_{inf(C) \cap R_D}\Big) \subseteq E.$$

Since E is deductively closed, we have that

$$inf(C)^* \cup max(C)^* \subseteq E.$$

Assume that $D_{min(C)} \subseteq GD(E, D)$. Then, we have that

$$Prereq\Big(D_{min(C)}\Big) \cup Conseq\Big(D_{min(C)}\Big) \subseteq E \qquad \text{and} \qquad min(C)^* \subseteq E$$

By Theorem 2, we have that

$$inf(C)^* \cup max(C)^* \cup min(C)^* \cup Prereq\Big(D_{max(C)}\Big)$$

is unsatisfiable. By monotonicity, this implies that E is unsatisfiable, a contradiction.

2. Analogous to 1.

$\blacksquare$

Proof 9 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 minimal conflicting set C in R. Let E be a consistent extension of (D_R, W).

1. Let $D_{max(C)} \subseteq GD(E, D)$. By the definition of $GD(E, D)$, this means that

$$Prereq\Big(D_{max(C)}\Big) \cup Conseq\Big(D_{max(C)}\Big) \subseteq E.$$

If $D_{min(C)} \subseteq GD(E, D)$ then we also have that

$$Prereq\Big(D_{min(C)}\Big) \cup Conseq\Big(D_{min(C)}\Big) \subseteq E.$$

Since $(min(C), max(C))$ is a conflicting core, we have

$$Prereq\Big(D_{max(C)}\Big) \cup Conseq\Big(D_{max(C)}\Big) \cup Prereq\Big(D_{min(C)}\Big) \cup Conseq\Big(D_{min(C)}\Big) \models \bot.$$

But this implies that E is unsatisfiable, a contradiction.

2. Follows analogously to 1.

∎

Proof 10 According to [Reiter, 1980], E is an extension of (D, W) iff $E = \bigcup_{i=0}^{\infty} E_i$, where

$$E_0 = W$$

and for $i \geq 0$

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

Transposing these definitions to **Z**-default theories of the form $(D_R, W \cup R_N^*)$ yields

$$E_0 = W \cup R_N^*$$

and (substituting general default rule $\tfrac{\alpha : \beta}{\omega}$ by the one in Definition 6):

$$E_{i+1} \;=\; Th(E_i) \;\cup\; \left\{ \beta_r \;\middle|\; \frac{\alpha_r \,:\, \beta_r \wedge \bigwedge_{r' \in R_r}(\alpha_{r'} \supset \beta_{r'})}{\beta_r} \in D_R, \right.$$
$$\left. \alpha_r \in E_i, \neg(\beta_r \wedge \bigwedge_{r' \in R_r}(\alpha_{r'} \supset \beta_{r'})) \notin E \right\},$$

where R_r is defined as in Definition 6). By relying on the fact that

$$D_R^n = \left\{ \tfrac{\alpha_r : \beta_r}{\beta_r} \;\middle|\; \alpha_r \to \beta_r \in R_D \right\},$$

the last but one equation can be easily transformed into

$$E_{i+1} \;=\; Th(E_i) \;\cup\; \left\{ \beta_r \;\middle|\; \tfrac{\alpha_r : \beta_r}{\beta_r} \in D_R^n,\; \alpha_r \in E_i, E \cup \{\beta_r\} \cup \bigcup_{r' \in R_r}(\alpha_{r'} \supset \beta_{r'}) \nvdash \bot \right\}$$

which completes our proof. ∎

Proof 11 The proof relies on the following two observations:

1. For fixed world knowledge W, if $\rho \in W$, then adding ρ to the justification of a default will not affect the resulting extensions.

   That is, for any default theory (D, W) if $\delta = \tfrac{\alpha : \beta}{\omega}$ and $\delta' = \tfrac{\alpha : \beta \wedge \rho}{\omega}$ then:

   E is an extension of $(D \cup \{\delta\}, W)$ and $\delta \in GD(E, D)$ iff E is an extension of $(D \cup \{\delta'\}, W)$ and $\delta' \in GD(E, D)$

2. For any $\alpha_r \Rightarrow \beta_r \in R_N$, $\alpha_r \supset \beta_r$ will be part of the world knowledge in the **Z**-default theory.

There are two cases to consider:

1. $\alpha_r \Rightarrow \beta_r \in R_N$, and $\alpha_r \Rightarrow \beta_r \in max(C)$ for conflicting core C.

   Let $\alpha_m \rightarrow \beta_m \in min(C)$ (or $\alpha_m \Rightarrow \beta_m \in min(C)$). So under the translation we will have a default of the form $\frac{\alpha_m : \beta_m \wedge \phi \wedge (\alpha_r \supset \beta_r)}{\beta_m}$. But by the second observation above, $\alpha_r \supset \beta_r$ is also in the world knowledge of the **Z**-default theory; by the first observation then, the default has precisely the same effect as $\frac{\alpha_m : \beta_m \wedge \phi}{\beta_m}$.

2. $\alpha_r \Rightarrow \beta_r \in R_N$, and $\alpha_r \Rightarrow \beta_r \in min(C)$ for conflicting core C.

   So under the translation we will have a default of the form $\frac{\alpha_r : \beta_r \wedge \phi}{\beta_r}$. But by the second observation above, $\alpha_r \supset \beta_r$ is also in the world knowledge of the **Z**-default theory. Hence the default $\frac{\alpha_r : \beta_r \wedge \phi}{\beta_r}$ is only trivially applicable in any extension (in that, if α_r is true, then since $\alpha_r \supset \beta_r$ is also true, we obtain that β_r is true, without reference to the default.)

The result follows by straightforward induction on $|R_N|$.

∎

Proof 12 Let $\langle R_D, R_N \rangle$ be a generic world description. Let us start with $R_D^{ab} \cup R_D^{sp} \cup R_N^*$. Consider

$$R_D^{sp} = \{\neg\bigwedge_{r' \in R_r}(\alpha_{r'} \supset \beta_{r'}) \supset ab_r \mid r \in R_D\}$$

where $R_r = \{r' \in max(C^i) \mid r \in min(C^i)$ for $i \in I \}$.

Since $(C^i)_{i \in I}$ is the family of all minimal conflicting sets in R this is equivalent to

$$R_D^{sp} = \{\neg\bigwedge_{r,r'}(\alpha_{r'} \supset \beta_{r'}) \supset ab_r \mid r' \in max(C^i) \text{ and } r \in min(C^i) \text{ for } i \in I\}$$

More transformations yield in turn:

$$R_D^{sp} = \{\bigwedge_{r,r'}(\alpha_{r'} \wedge \neg ab_r \supset \beta_{r'}) \mid r' \in max(C^i) \text{ and } r \in min(C^i) \text{ for } i \in I\}$$

$$R_D^{sp} = \{\bigwedge_{r \in \overline{R}_{r'}}(\alpha_{r'} \wedge \neg ab_r \supset \beta_{r'}) \mid r' \in R_D\}$$

$$\text{where } \overline{R}_{r'} = \{r \in min(C^i) \mid r' \in max(C^i) \text{ for } i \in I \}.$$

Switching r and r' gives us:

$$R_D^{sp} = \{\alpha_r \wedge \left(\bigvee_{r' \in R_r'} \neg ab_{r'}\right) \supset \beta_r \mid r \in R_D\}$$

$$\text{where } R'_r = \{r' \in min(C^i) \mid r \in max(C^i) \text{ for } i \in I \}.$$

Recall that

$$R_D^{ab} = \{\alpha_r \wedge \neg ab_r \supset \beta_r \mid r \in R_D\}.$$

Consequently, we may combine all elements in R_D^{ab} with those in the latest equation describing R_D^{sp}, which gives us:

$$R_D^\star = \{\alpha_r \wedge \left(\neg ab_r \vee \bigvee\nolimits_{r' \in R'_r} \neg ab_{r'}\right) \supset \beta_r \mid r \in R_D\}$$

where $R'_r = \{r' \in min(C^i) \mid r \in max(C^i) \text{ for } i \in I \}$.

As a consequence, we have shown that $R_D^{ab} \cup R_D^{sp} \cup R_N^*$ is transformable into $R_D^\star \cup R_N^*$ by consecutive application of logical and set-theoretic transformations.
∎

Proof 13 Let R be a set of rules and let W be a set of formulas. Let C be a minimal conflicting set in R.

1. Let M be a model of

   $$Circum(W \cup R_D^\star \cup R_N^*; \{ab_r \mid r \in R_D\}; Z) \cup inf(C)^* \cup \{\alpha_r \wedge \neg ab_r \mid r \in max(C)\} .$$

   Then, M is also a model of $\{\beta_r \mid r \in max(C)\}$; and therefore it also entails $max(C)^*$.

   Now, assume M is also a model of

   $$\{\beta_r \mid r \in min(C)\}.$$

   Then, M is also a model of $min(C)^*$.

   By Theorem 2, we have however that

   $$inf(C)^* \cup max(C)^* \cup min(C)^* \cup \{\alpha_r \mid r \in max(C)\}$$

   is unsatisfiable, a contradiction to the fact that M is a model!

2. Analogous to 1.

∎

Proof 14 Let R be a set of rules and let W be a set of formulas. Let C be a minimal conflicting set in R.

1. Let M be a model of

$$Circum(W \cup R_D^\star \cup R_N^*; \{ab_r \mid r \in R_D\}; Z) \cup \{\alpha_r \wedge \neg ab_r \mid r \in max(C)\}\ .$$

Then, M is also a model of $\{\beta_r \mid r \in max(C)\}$; and therefore it also entails $\{\alpha_r \wedge \beta_r \mid r \in max(C)\}$.

Now, assume M is also a model of

$$\{\alpha_r \wedge \beta_r \mid r \in min(C)\}.$$

Since $(min(C), max(C))$ is a conflicting core, we have

By Theorem 2, we have however that

$$\{\alpha_r \wedge \beta_r \mid r \in min(C)\}. \cup \{\alpha_r \wedge \beta_r \mid r \in max(C)\}$$

is unsatisfiable, a contradiction to the fact that M is a model!

2. Analogous to 1.

$\blacksquare$

Proof 15 Let M be a model of $R_N^* \cup R_D^*$ such that $M \cap \{ab_r \mid r \in R_D\} = \emptyset$. That is, M is a model falsifying all ab_r. Clearly, such a model exists, since there is no occurrence of any ab_r in $R_N^* \cup R_D^*$. By definition,

$$M \models \alpha_r \supset \beta_r \qquad \text{for all } r \in R_D.$$

This implies that

$$M \models \alpha_r \wedge \left(\neg ab_r \vee \bigvee_{r' \in R_r'} \neg ab_{r'} \right) \supset \beta_r \qquad \text{for all } r \in R_D$$

because $M \models \neg ab_r$. Hence, M is also a model of $R_D^\star$. This extends clearly to $R_N^* \cup R_D^\star$ since M is a model of R_N^* by definition. $\blacksquare$

Proof 16 Let M be a model of $W \cup R_N^*$. Consider $M' = M \cup \{ab_r \mid r \in R_D\}$. Clearly, $M' \models W \cup R_N^*$, since there is no occurrence of any ab_r in $W \cup R_N^*$. Moreover, we have

$$M' \models \alpha_r \wedge \left(\neg ab_r \vee \bigvee_{r' \in R_r'} \neg ab_{r'} \right) \supset \beta_r \qquad \text{for all } r \in R_D^\star$$

because $M' \models ab_r$ for all $r \in R_D$.

Since

$$M' \models W \cup R_N^* \cup R_D^*$$

we can essentially begin with M' and eliminate (non-deterministically) one atom ab_r after another as long as the resulting interpretation is a model of $W \cup R_N^* \cup R_D^*$. While varying our initial model M, we obtain in this way all minimal models of $W \cup R_N^* \cup R_D^*$. Such models exist since we deal with a propositional language over a finite alphabet, so that there is only a finite number of atoms like ab_r. This shows that the circumscription results in a satisfiable formula. ∎

Acknowledgements

References

[Adams, 1975] E. Adams. *The Logic of Conditionals*. D. Reidel Publishing Co., Dordrecht, Holland, 1975.

[Baader and Hollunder, 1993a] F. Baader and B. Hollunder. How to prefer more specific defaults in terminological default logic. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 669–674, Chambéry, France, 1993.

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

[Benferhat *et al.*, 1992] S. Benferhat, D. Dubois, and H. Prade. Representing default rules in possibilistic logic. In *Proceedings of the Third International Conference on the Principles of Knowledge Representation and Reasoning*, pages 673–684, Cambridge, MA, October 1992.

[Boutilier, 1992a] C. Boutilier. *Conditional Logics for Default Reasoning and Belief Revision*. PhD thesis, Department of Computer Science, University of Toronto, 1992.

[Boutilier, 1992b] C. Boutilier. What is a default priority? In J. Glasgow and R. Hadley, editors, *Proceedings of the Ninth Canadian Conference on Artificial Intelligence*, pages 140–147, Vancouver, B.C., 1992.

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

[Brewka, 1993] G. Brewka. Adding priorities and specificity to default logic. Technical report, GMD, 1993.

[Chellas, 1980] B.F. Chellas. *Modal Logic*. Cambridge University Press, 1980.

[Clark, 1978] K.L. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Databases*, pages 293–322. Plenum Press, 1978.

[Delgrande and Schaub, 1994a] J.P. Delgrande and T. Schaub. A general approach to specificity in default reasoning. In J. Doyle, P. Torasso, and E. Sandewall, editors, *Proceedings of the Fourth International Conference on the Principles of Knowledge Representation and Reasoning*, pages 146 – 157, Bonn, Germany, May 1994.

[Delgrande and Schaub, 1994b] J.P. Delgrande and T. Schaub. Incorporating specificity into circumscriptive theories. In B. Nebel and L. Dreschler-Fischer, editors, *KI'94: Advances in Artificial Intelligence, Proceedings of the Eighteenth German Annual Conference on Artificial Intelligence*, volume 861 of *Lecture Notes in Artificial Intelligence*, pages 272–283. Springer-Verlag, 1994.

[Delgrande *et al.*, 1994] J.P. Delgrande, T. Schaub, and W.K. Jackson. Alternative approaches to default logic. *Artificial Intelligence*, 70:167–237, 1994.

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

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

[Dubois and Prade, 1991] D. Dubois and H. Prade. Conditional objects and non-monotonic reasoning. In *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning*, pages 175–185, Cambridge, MA, April 1991.

[Eiter and Gottlob, 1993] T. Eiter and G. Gottlob. Propositional circumscription and extended closed world reasoning are Π_2^p-complete. *Theoretical Computer Science*, 114:231–245, 1993.

[Etherington and Reiter, 1983] D.W. Etherington and R. Reiter. On inheritance hierarchies with exceptions. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 104–108. Morgan Kaufmann Publishers, 1983.

[Fahlman *et al.*, 1981] S.E. Fahlman, D.S. Touretzky, and W. van Roggen. Cancellation in a parallel semantic network. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 257–263, Vancouver, B.C., 1981.

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

[Ginsberg, 1989] M. Ginsberg. A circumscriprive theorem prover. *Artificial Intelligence*, 39:209–230, 1989.

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

[Goldszmidt, 1992] M. Goldszmidt. *Qualitative Probabilities: A Normative Framework for Commonsense Reasoning*. PhD thesis, Department of Computer Science, University of California, Los Angeles, 1992.

[Grosof, 1991] B. Grosof. Generalizing prioritization. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning*, pages 289–300, San Mateo, CA, April 1991. Morgan Kaufmann.

[Horty, 1994] J.F. Horty. Some direct theories of nonmonotonic inheritance. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Nonmonotonic Reasoning and Uncertain Reasoning*, volume 3 of *Handbook of Logic in Artifical Intelligence and Logic Programming*, pages 111–187. Oxford: Clarendon Press, 1994.

[Hughes and Cresswell, 1968] G.E. Hughes and M.J. Cresswell. *An Introduction to Modal Logic*. Methuen and Co. Ltd., 1968.

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

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

[Lamarre, 1991] Philippe Lamarre. S4 as the conditional logic of nonmonotonicity. In *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning*, pages 357–367, Cambridge, MA, April 1991.

[Lehmann, 1989] D. Lehmann. What does a conditional knowledge base entail? In R. Brachman, H. Levesque, and R. Reiter, editors, *Proceedings of the First International Conference on the Principles of Knowledge Representation and Reasoning*, pages 212–222, Toronto, Canada, May 1989.

[Lewis, 1973] D. Lewis. *Counterfactuals*. Harvard University Press, 1973.

[Lifschitz, 1985] V. Lifschitz. Computing circumscription. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 121–127, Los Angeles, 1985.

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

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

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

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

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, CA, 1988.

[Pearl, 1989] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proceedings of the First International Conference on the Principles of Knowledge Representation and Reasoning*, pages 505–516, Toronto, May 1989. Morgan Kaufman.

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

[Poole, 1985] D. Poole. On the comparison of theories: Preferring the most specific explanation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 144–147, 1985.

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

[Reiter and Criscuolo, 1981] R. Reiter and G. Criscuolo. On interacting defaults. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 270–276, Vancouver, B.C., 1981.

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

[Shoham, 1988] Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. The MIT Press, Cambridge, MA, 1988.

[Stalnaker, 1968] R.C. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell, Oxford, 1968.

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