Some Direct Theories of Nonmonotonic Inheritance

John F. Horty

Contents

1 Introduction

1.1 Background

Inheritance systems were originally developed within artificial intelligence in response to the practical need for an efficient way of representing and accessing taxonomic information. These systems, along with network representations more generally, were first presented without any semantic analysis at all, or else only with a procedural semantics, according to which the meaning of the representations was supposed to be specified implicitly by the programs operating on them. However, in large part due to the criticisms of [Woods, 1975], a good deal of attention was soon drawn to the problem of supplying an implementation independent account of the meaning of these network formalisms. Most of the work on this problem, and notably that of [Hayes, 1979], followed a translational or *indirect* strategy, specifying the meaning of a network formalism through a mapping into an ordinary logical language, usually classical first order logic. Because of these mappings, it was generally concluded that the networks could be regarded simply as notational variants of syntactically restricted first order theories—distinguished, perhaps, by supporting specialized inference algorithms, or perhaps, as [Hayes, 1977] suggests, only by their attractive appearance on the printed page.

With the attempt to incorporate defeasible information into inheritance hierarchies—in systems such as FRL [Roberts and Goldstein, 1977], KRL [Bobrow and Winograd, 1977], and NETL [Fahlman, 1979]—questions concerning the precise meaning of these network representations arose once again, and with a new urgency, since it had been shown in [Fahlman *et al.*, 1981], for example, that the naive inference algorithms built into these systems could lead to bizarre and unintuitive results. Because their informal interpretation required a nonmonotonic consequence relation, it was plain that that these representational formalisms could not naturally be translated into classical logic. Nevertheless, it seemed to many that the indirect approach could be extended also to this case by translating the networks into one or another of the nonmonotonic logics. The first project along these lines was the interpretation in [Etherington and Reiter, 1983] of simplified NETL-style networks in terms of default logic. In analogy with the earlier work of Hayes and others, this research was taken to support the conclusion that defeasible inheritance networks could be viewed as syntactically restricted default theories.

At approximately the same time, however, a very different kind of technique for analyzing the meaning of defeasible inheritance networks was being developed—initially, in a 1984 dissertation by Touretzky, published two years later as [Touretzky, 1986]. The point of providing a semantic theory for some representational formalism is to allow us to delineate the consequences of a set of facts expressed in that formalism, to explore the characteristics of these consequence sets, and to test the original facts for properties such as consistency. One way to do this, of course, is by mapping the representational formalism into some logical language for which ideas like consequence and consistency have already been defined. But as Touretzky noticed, a theory that achieves the same ends can also be developed, at least in the case of inheritance networks, entirely in terms of the network language itself, without going through the intermediate step of translation into a separate logic. By analyzing a number of examples, it is possible to develop detailed intuitions about the kind of conclusions that should be derivable from inheritance networks. These intuitions can be codified into a general and mathematically rigorous "inheritance definition," which can itself be tested against further examples, and used as a basis for establishing semantic metatheorems. In contrast to the translational approach, an account of the meaning of inheritance networks that proceeds along these lines can be described as non-translational or *direct*.

These direct theories of inheritance occupy a somewhat ambiguous position with regard to the usual contrast between declarative and procedural methods. For theories of this kind, the paths through a network often form the main focus of attention (because of this, direct theories are sometimes described as *path-based*). These paths, which are analogous to proofs in ordinary logic, are supposed to represent arguments or inference procedures; and they do tend to correspond in a loose way to the procedures actually carried out in implementations of inheritance systems. To that extent, then, the approach is explicitly procedural. In contrast to the usual conception of procedural semantics, however, the direct theories of inheritance are not dependent upon any particular implementation. Instead, they aim to achieve the declarative goal of providing an independent standard against which implementations are to be judged.

Another way in which these direct theories differ from the standard declarative paradigm is that they tend not to rely on model theoretic notions, which were taken by some researchers [Hayes, 1977; McDermott, 1978] to form the core of any rigorous semantics for a representational formalism. The account of meaning provided in these direct theories for a statement belonging to the network language tells us what can be derived from a network containing that statement, as well as the conditions under which that statement itself can be derived from some network. The treatment of derivability is purely syntactic (or proof theoretic); it does not rely on a prior notion of truth in a structure for items belonging to the network language. In fact, there is no attempt at all to define truth conditions for defeasible links occurring in inheritance networks—and the direct approach has sometimes been criticized for this reason.

This kind of criticism seems to be misguided; or at least, it relies on standards very different from those at work in other areas of logic, where proof theory and model theory stand on a more even footing. Historically, at least, proof theory came first. The proof theory for both classical and intuitionistic logic was already well understood by the time formal model theoretic techniques were introduced. Many of the fundamental metatheoretic results for intuitionistic logic in particular were first established by proof theoretic means. As shown by the articles of [Barwise, 1977, Part D], for example, proof theory remains an active area of research, especially for those concerned with constructivity. And there are logicians, such as Prawitz [Prawitz, 1971; Prawitz, 1972], who feel for broadly philosophical reasons that a proof theoretic analysis of logical concepts is actually to be preferred to a model theoretic approach.

A more balanced view would take proof theoretic and model theoretic approaches as providing different but largely complementary kinds of insights into logical notions. In many cases, such as first order classical logic and most of the familiar nonstandard logics, the two approaches are linked through completeness results. However, there are situations in which the approaches fall apart, and one or the other seems to be more fruitful. Often, for example, we are faced with logics for which the semantic consequence relation is not axiomatizable, and of course, standard proof theory is of little use here. On the other hand, proof theoretic techniques seem to be especially illuminating for the study of logics in which the meanings assigned to connectives are closely intertwined with the structure of argumentation. The family of relevance logics [Anderson and Belnap, 1975] provides a good example: although many of these logics were eventually supplied with a model theory of the possible worlds variety, most people do not find this to be as useful or illuminating for understanding these systems as their original proof theoretic presentation.

The direct theories of inheritance described here are best thought of as analogous to proof theoretic work in ordinary logic; they aim to provide a precise account of correct reasoning in defeasible inheritance networks that does not appeal to model theoretic techniques. Of course, the resulting account is significantly different from what is found in ordinary proof theory, particularly because it is forced to deal with interactions among arguments, rather than single arguments in isolation. But this should not be too surprising: nonmonotonic logic is significantly different from ordinary logic, and at least the fixed point approaches to nonmonotonicity concentrate on statement sets rather than single statements.

It would be nice if this proof theoretic approach to defeasible inheritance could be supplemented with a model theoretic treatment, and a careful study of the proof theory may suggest ways in which classical model theory could be adapted to the new situation. In fact, this sort of thing has already happened in the case of inheritance networks without defeasible information, where the direct analysis has suggested that a certain nonstandard model theory should be employed (see Section 3.1 below). But the introduction of defeasible information changes matters. Just as with relevance logic, the direct theories of defeasible inheritance are closely bound up with the structure of argumentation. They rely on the kind of detailed syntactic concerns that are difficult to work into a model theory; and even if this could be done, it is hard to tell whether the resulting account would yield a perspective significantly different from that already provided by the proof theoretic approach.

1.2 Overview

The past few years have seen intense activity in the semantic analysis of nonmonotonic inheritance systems. A number of different accounts, both direct and translational, have been developed within a variety of theoretical frameworks, and it would be impossible to treat them all with any care. Instead, although we try to relate the work described here to other research in the vicinity, this chapter concentrates only on a few theories developed within a single framework—roughly, those based on Touretzky's original ideas or their close relatives. This is the work that has been most influential in shaping subsequent research. Most of the theories described here have appeared elsewhere in the literature, or are easily assembled from ideas appearing elsewhere. The goal of the chapter, however, is to present this work—which is scattered throughout a number of publications—in a uniform notation and from a uniform point of view; and this has forced a certain amount of reformulation and restructuring of ideas. The entire chapter should be accessible to a reader new to the area, but even someone already familiar with the work may find the uniform presentation helpful.

The chapter is organized as follows. Section 2 presents some direct theories of inheritance for purely defeasible networks; Section 3 shows how these theories can be generalized to apply also to networks containing strict and defeasible links mixed together; Section 4 describes some variations on these basic theories and explores some of the problems they suggest; and an appendix proves some sample theorems.

Because the volume in which this chapter appears is devoted to nonmonotonic reasoning in general, rather than the development of knowledge representation technologies, inheritance theories are presented here with a particular bias: conceptual issues regarding correctness and intuitive motivation are emphasized at the expense of algorithmic and implementational considerations. A more balanced survey, which places more emphasis on the applications of inheritance in knowledge representation is found in [Thomason, 1992].

From the general standpoint of nonmonotonic reasoning, the direct theories of inheritance described here seem to be of interest primarily for two reasons. First, because defeasible inheritance networks provide such an easy way of visualizing complicated patterns of interaction among defaults, they form a good problem domain for testing ideas formulated in more general nonmonotonic logics; and they are often used for this purpose. For people testing their ideas in this way, by formalizing inheritance networks, the direct, nontranslational approaches can provide a sort of "pretheoretic" indication of the results they might wish to achieve. These theories can then be thought of as a bridge discipline standing between the applications of inheritance networks in knowledge representation and their formalization in more generally applicable nonmonotonic logics.

Second, although the matter is contentious, it can be argued that the direct theories described here have, to date, actually been more successful at capturing the intuitive meaning of defeasible inheritance networks than their translational counterparts. The reason for this seems to be that the explicit representation of arguments in these direct theories allows for a very fine-grained analysis of the structure of defeasible reasoning, which it is often difficult to achieve using more general nonmonotonic logics. For this reason also, because of their sensitivity to the detailed structure of argument, the techniques of path-based inheritance allow a good deal of versatility in the definition of particular theories, enabling us to articulate a variety of different intuitions about defeasible reasoning. At times, the contrasts among these different theories can suggest, in a very simple environment, both options and problems that may not have been apparent from a more general point of view.

1.3 Basic concepts

1.3.1 Links and paths

The kind of inheritance networks considered here, as in most of the theoretical literature, are simple collections of positive and negative IS-A links—very broad idealizations of the systems actually used in knowledge representation. In our description of these networks, letters from the beginning of the alphabet (a through d) refer to objects or individuals; letters from the middle of the alphabet (m through t) refer to properties or kinds; and letters from the end of the alphabet (u through z) range over both objects and properties.

The link types $x \Rightarrow p$ and $x \not\Leftrightarrow p$ represent positive and negative *strict* statements. If x is a property, these positive and negative strict links are equivalent to quantified conditionals: the link $q \Rightarrow p$ represents a statement of the form 'Every Q is a P'; the link $q \not\Leftrightarrow p$ represents a statement of the form 'No Q is a P'. If x is an object, these strict links are equivalent to ordinary literals: $a \Rightarrow p$ and $a \not\Leftrightarrow p$ represent the statements Pa and $\neg Pa$. The link types $x \rightarrow p$ and $x \not\rightarrow p$ represent positive and negative *defeasible* statements. If x is a property, these defeasible links correspond to generic statements: $q \rightarrow p$ and $r \not\rightarrow p$, for example, might stand for the statements 'Birds fly' and 'Mammals don't fly'. There is nothing in classical logic very close in meaning to statements like these. For example, 'Birds fly' does not mean that all birds fly, since it is true even in the presence of exceptions. Instead, it seems to mean that "typical birds" fly—or that, for any given bird a, it is most natural to suppose that a flies. If x is an object, it is more difficult to find a simple reading for these defeasible links; but we will assume that $a \rightarrow p$ and $a \not\rightarrow p$ mean something along the lines of 'It is most natural to suppose that Pa' and 'It is most natural to suppose that $\neg Pa$'.

Lower case Greek letters range over *paths*, to be defined as special sequences of links. Often, it is convenient to refer to an arbitrary path in a way that displays some of the nodes it passes through without displaying the particular link types connecting those nodes. For this purpose, we adopt a notation according to which $\pi(x, \sigma, y)$ represents an arbitrary positive path, and $\overline{\pi}(x, \sigma, y)$ likewise to an arbitrary negative path, from x through σ to y. As a convention governing this π-notation, we assume that adjacency of node symbols entails adjacency of nodes on the paths symbolized. Thus, for example, $\overline{\pi}(x, u, \sigma, y)$ represents a negative path beginning with a direct link of any type from x to u, and then moving through σ to y.

Paths are classified as simple or compound, strict or defeasible, positive or negative. The simple paths are just the direct links—classified as strict or defeasible, positive or negative, along with the links themselves. The compound paths are defined inductively, as follows:

1. If $\pi(x, \sigma, p)$ is a strict positive path, then: $\pi(x, \sigma, p) \Rightarrow q$ is a strict positive compound path, $\pi(x, \sigma, p) \not\Leftrightarrow q$ is a strict negative compound path, $\pi(x, \sigma, p) \rightarrow q$ is a defeasible positive compound path, and $\pi(x, \sigma, p) \not\rightarrow q$ is a defeasible negative compound path;

2. If $\overline{\pi}(x, \sigma, p)$ is a strict negative path, then: $\pi(x, \sigma, p) \Leftarrow q$ is a strict negative compound path;

3. If $\pi(x, \sigma, p)$ is a defeasible positive path, then: $\pi(x, \sigma, p) \Rightarrow q$ is a defeasible positive compound path, $\pi(x, \sigma, p) \not\Leftrightarrow q$ is a defeasible negative compound path, $\pi(x, \sigma, p) \rightarrow q$ is a defeasible positive compound path, and $\pi(x, \sigma, p) \not\rightarrow q$ is a defeasible negative compound path;

4. If $\overline{\pi}(x, \sigma, p)$ is a defeasible negative path, then: $\pi(x, \sigma, p) \Leftarrow q$ is a defeasible negative compound path.

Intuitively, paths represent arguments, which *support* certain statements as their conclusions. A positive path of the form $\pi(x, \sigma, y)$ supports the statement $x \Rightarrow y$ if it is strict and the statement $x \rightarrow y$ if it is defeasible; likewise, a negative path of the form $\overline{\pi}(x, \sigma, y)$ supports $x \not\Leftrightarrow y$ if it is strict and $x \not\rightarrow y$ if it is defeasible.

1.3.2 Nets, theories, and extensions

Capital Greek letters from the beginning of the alphabet $(\Gamma, \Delta, \Theta, \ldots)$ stand for *networks* (or simply: nets), which are finite sets of links. A network is *defeasible* if it contains only defeasible links, *strict* if it contains only strict links, and *mixed* if it contains both strict and defeasible links. Capital Greek letters from the end of the alphabet $(\Phi, \Xi, \Psi, \ldots)$ stand for sets of paths in general. Intuitively, the statements belonging to a network are supposed to represent the information provided as hypotheses to some agent or reasoning mechanism; the members of a path set are supposed to represent those patterns of reasoning that have been explicitly carried out and accepted. The relation of support already defined between paths and statements can be extended in the obvious way to a relation between path sets and statements sets: a path set Φ will be said to *support* a statement set Δ just in case Δ is the set of statements supported by the paths in Φ.

The primary goal of the semantic account of inheritance networks is to specify the *theories* associated with each network—the statement sets that an ideal reasoner could arrive at, given the information in that network as hypotheses, or initial information. Rather than attempting to define this relation between nets and their theories directly, however, we follow a roundabout route. We first specify a relation between networks and certain path sets, known as their *extensions*. Intuitively, an extension of a net represents some total set of argument paths that an ideal reasoner might accept, based on the initial information in that net. Once this relation has been defined, it is then a simple matter to specify the theories associated with a network: Δ is a theory of the network Γ just in case there is an extension Φ of Γ such that Φ supports Δ.

1.3.3 Contexts and inheritability

If Γ is a network and Φ is some set of paths, we will describe the pair $\langle \Gamma, \Phi \rangle$ as an *epistemic context*. Although, formally, any such pairing of a net and a path set counts as an context, it is part of the intuitive picture that the path set should arise out of the net: we imagine that an agent is provided with Γ as his initial information, and that after a certain amount of reasoning based on this information, he has been led to accept the set Φ of arguments.

In any given context, certain arguments or paths can be classified as *inheritable*—forcible or persuasive, in that context. We will use the symbol $\mid\!\sim$ to represent this relation of inheritability, so that $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$ means that the path σ is inheritable in the context $\langle \Gamma, \Phi \rangle$. This notion of inheritability is the central concept in our treatment of semantic networks; it plays a crucial role in our characterization of extensions.

2 Theories of defeasible inheritance

This section develops three theories of inheritance for defeasible networks—a theory modeled on Touretzky's original credulous approach, a skeptical theory, and then a flexible theory that slips between the other two. The development takes place in two stages: after setting out a notion of inheritability appropriate for defeasible nets, we use this notion to define the different kinds of extensions characteristic of the three theories.

Although these theories yield different results as the appropriate conclusions of a network, it is not necessary to view them as competitors, in the sense that one is right and the others are wrong. It is possible instead to imagine that the different theories are appropriate in different kinds of reasoning situations.

2.1 Defeasible inheritability

The notion of inheritability presented here is modeled on that of [Touretzky, 1986]; it relies on three preliminary concepts, which we turn to first—constructibility, conflict, and preemption.

The paths constructible in an epistemic context are those that can be assembled by chaining together, in a certain way, the paths and links already present in that context.

Definition 2.1.1 (Constructibility). A positive path $\pi(x, \sigma, u) \rightarrow y$ is *constructible* in the context $\langle \Gamma, \Phi \rangle$ iff $\pi(x, \sigma, u) \in \Phi$ and $u \rightarrow y \in \Gamma$. A negative path $\pi(x, \sigma, u) \not\rightarrow y$ is *constructible* in the context $\langle \Gamma, \Phi \rangle$ iff $\pi(x, \sigma, u) \in \Phi$ and $u \not\rightarrow y \in \Gamma$.

Each path represents an argument, and it is useful to think of the final link in a constructible path as a *reason* for accepting that argument. For example, suppose an agent is provided with the simple network $\Gamma = \{a \rightarrow p, p \rightarrow q, q \rightarrow r\}$ as his initial information, where a = Tweety, p = canaries, q = birds, and r = flying things; and suppose also that he has not yet drawn any inferences from this information, so that his epistemic context is just $\langle \Gamma, \Gamma \rangle$. In this context, the argument path $a \rightarrow p \rightarrow q$ is constructible: the agent is provided with the statement that Tweety is a canary as part of his initial information, and so the statement that canaries are birds provides him with a reason for accepting the argument that Tweety is a bird. Now imagine that the agent actually does accept this argument, so that he moves to the new context $\langle \Gamma, \Phi \rangle$, where $\Phi = \Gamma \cup \{a \rightarrow p \rightarrow q\}$. In this new context, the path $a \rightarrow p \rightarrow q \rightarrow r$ will be constructible as well: since he has already accepted the argument that Tweety is a bird, the information that birds fly gives him a reason to accept the further argument that Tweety flies.

Constructibility is a necessary condition that a compound path must satisfy in order to be classified as inheritable, but it is not sufficient. Even if a context provides some reason for accepting an argument, that argument might not, all things considered, count as forcible or persuasive. Two further aspects of the context could interfere.

First, an argument will not be classified as inheritable in a context if it is conflicted, where this notion is defined as follows.

Definition 2.1.2 (Defeasible conflict). A path of the form $\pi(x, \sigma, y)$

conflicts with any path of the form $\overline{\pi}(x, \tau, y)$.

Definition 2.1.3 (Conflicted paths). A path σ is *conflicted* in the context $\langle\Gamma, \Phi\rangle$ iff Φ contains a path that conflicts (in Γ) with σ.[1]

The intuitive force of this restriction is that the agent should check for consistency before accepting arguments: a defeasible argument cannot be classified as forcible, even if there is some reason for accepting it, whenever its adoption would introduce a conflict into an epistemic context. The kind of situation in which this restriction comes into play is best illustrated by a standard example. Suppose, then, that the agent is provided with the net Γ_1 (Figure 1, known as the Nixon Diamond) as initial information, where a = Nixon, q = Quakers, r = republicans, and p = pacifists; and again, that he has not yet drawn any additional conclusions, so that his epistemic context is simply $\langle\Gamma_1, \Gamma_1\rangle$. In that case, both of the constructible paths $a \rightarrow q \rightarrow p$ and $a \rightarrow r \not\rightarrow p$ should be classified as inheritable. Although these arguments conflict with each other, neither is conflicted in the context; both represent forcible arguments, and so the agent has to find a way of dealing with the conflicting implications of his current epistemic state. Now let us imagine that, out of some motive, the agent actually adopts one of these two arguments—say, the argument $a \rightarrow r \not\rightarrow p$—so that he moves to the new context $\langle\Gamma_1, \Phi\rangle$, where $\Phi = \Gamma_1 \cup \{a \rightarrow r \not\rightarrow p\}$. In this new context, the path $a \rightarrow q \rightarrow p$ is conflicted and so no longer inheritable; the argument loses its force for the agent, since he has already decided to accept an argument to the contrary.

The second restriction governing inheritability is that a constructible path cannot be classified as inheritable in a context if it is preempted. This restriction is supposed to reflect the idea that an agent should not view an argument as persuasive, even if he has some reason for accepting it, whenever his epistemic context provides a better reason for accepting a conflicting argument. Again, it is best to illustrate this idea with a familiar example; so suppose that Γ_2 (Figure 2, known as the Tweety Triangle) represents the agent's initial information, where a = Tweety r = penguins, q = birds, and p = flying things. Let us assume that the agent has already concluded that Tweety is a bird—that he has reasoned his way to the epistemic context $\langle\Gamma_2, \Phi\rangle$, where $\Phi = \Gamma_2 \cup \{a \rightarrow r \rightarrow q\}$. In this context, the paths $a \rightarrow r \rightarrow q \rightarrow p$ and $a \rightarrow r \not\rightarrow p$ are both constructible: the link $q \rightarrow p$ provides a reason for the conclusion that Tweety flies; the link $r \not\rightarrow p$ provides a reason for the conclusion that Tweety does not fly. As in the case of Γ_2, these paths conflict with each other, though neither is conflicted in the context. Nevertheless, it does not seem on intuitive grounds that

[1] In the following section on mixed networks, the notion of conflict between paths is replaced by a notion of conflict between paths relative to a network; the parenthetical clause in this definition anticipates the development.

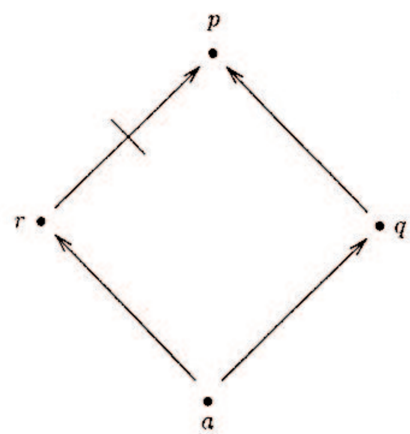

Fig. 1. Γ_1

the agent should be free, as before, to adopt whichever one of these two paths he chooses; here the first seems to be, in some sense, undermined or preempted by the second.

It is important to see exactly what it is about the context $\langle\Gamma_2, \Phi\rangle$, that leads us to view $a \rightarrow r \rightarrow q \rightarrow p$ as a preempted path—what it is that makes us dismiss the link $q \rightarrow p$ as a persuasive reason for accepting this argument. There seem to be two components to the notion. First, Φ contains a path *from a through r to q*, so that conclusions about a based on r can be thought of as "more specific," or in some other way better, than conclusions about a based on q; and second, the direct link $r \not\rightarrow p$ belongs to Γ_2, so that r can be thought of as providing "immediate" information contrary to that provided by q. Generalizing these observations leads to the following definition.

Definition 2.1.4 (Defeasible preemption). A positive path $\pi(x, \sigma, u) \rightarrow y$ is *preempted* in the context $\langle\Gamma, \Phi\rangle$ iff there is a node v such that (i) either $v = x$ or there is a path of the form $\pi(x, \tau_1, v, \tau_2, u) \in \Phi$, and (ii) $v \not\rightarrow y \in \Gamma$. A negative path $\pi(x, \sigma, u) \not\rightarrow y$ is *preempted* in the context $\langle\Gamma, \Phi\rangle$ iff there is a node v such that (i) either $v = x$ or there is a path of the form $\pi(x, \tau_1, v, \tau_2, u) \in \Phi$, and (ii) $v \rightarrow y \in \Gamma$.

When a positive path of the form $\pi(x, \sigma, u) \rightarrow y$ is preempted in accord with this definition (and similarly for negative paths), we say that it is preempted

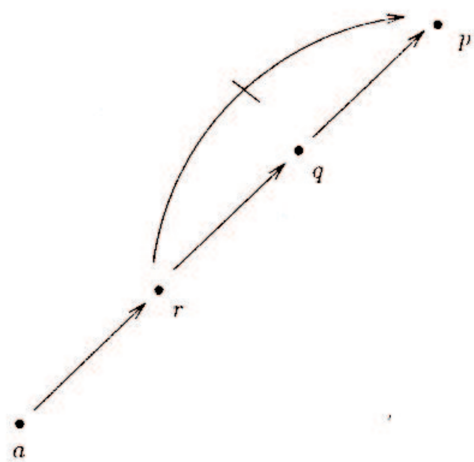

Fig. 2. Γ_2

by the path $\pi(x, \tau_1, v) \rightarrow y$, and we refer to this path as a *preemptor*. The path $\pi(x, \tau_1, v, \tau_2, u)$ is referred to as a *situator*, since it situates the node v appropriately between x and u; and the node v itself, which acts as the fulcrum for preemption, is referred to as the *preempting node*. In the case of Γ_2, for example, $a \rightarrow r \not\rightarrow p$ is the preemptor, $a \rightarrow r \rightarrow q$ is the situator, and r is the preempting node.

At this point, we can assemble our preliminary concepts into a formal definition of inheritability for defeasible networks.

Definition 2.1.5 (Defeasible inheritability).

Case I: σ is a direct link. Then $\langle \Gamma, \Phi \rangle \mathrel{|\!\!\sim} \sigma$ iff $\sigma \in \Gamma$.

Case II: σ is a compound path. Then $\langle \Gamma, \Phi \rangle \mathrel{|\!\!\sim} \sigma$ iff

1. σ is constructible in $\langle \Gamma, \Phi \rangle$,
2. σ is not conflicted in $\langle \Gamma, \Phi \rangle$,
3. σ is not preempted in $\langle \Gamma, \Phi \rangle$.

This definition differs from that of [Touretzky, 1986] only in the following details: first, the present definition allows direct links to be classified as inheritable; second, it relies on an alternative notion of constructibility for compound inheritable paths; and third, it relies on an alternative notion of preemption. The first of these differences is inessential, simply allowing a more straightforward definition of the extensions. The other two differences will be discussed later, in Sections 4.1 and 4.2.

2.2 Constructing extensions

Now that we have defined a relation of defeasible inheritability, we use can this notion to characterize the different kinds of extensions.

2.2.1 Credulous extensions

Perhaps the most natural class of extensions, and certainly the simplest to motivate, are the credulous extensions, due to Touretzky. Intuitively, an extension is supposed to represent some total set of arguments that an ideal reasoner would be able to accept, based on the initial information in a network. To motivate the credulous extensions, then, we need only ask: what could prevent some path set Φ from representing such an ideal set of arguments determined by the net Γ? There are two obvious possibilities. First, Φ might contain too few arguments; there might be some argument inheritable in the context $\langle\Gamma, \Phi\rangle$ that does not actually belong to Φ. Second, Φ might contain too many arguments; some argument actually belonging to Φ might turn out not to be inheritable in the context. The credulous extensions of a net Γ can be defined as those path sets exhibiting neither of these defects.

Definition 2.2.1. The path set Φ is a *credulous extension* of the net Γ iff

$$\Phi = \{\sigma : \langle\Gamma, \Phi\rangle \mathrel{|\!\!\sim} \sigma\}.$$

Again, this notion is essentially that of [Touretzky, 1986]; the present formulation simply reorganizes the component ideas a bit, so that the credulous extensions can be characterized explicitly as fixed points based on the inheritability relation.[2]

Of course, a given semantic net may possess more than one credulous extension; for example, the path sets $\Gamma_1 \cup \{a \rightarrow q \rightarrow p\}$ and $\Gamma_1 \cup \{a \rightarrow r \not\rightarrow p\}$ are both credulous extensions of the net Γ_1. In such a case, when a net possesses more than one credulous extension, it is not clear exactly how to characterize the set of conclusions that an ideal reasoner should draw from the information contained in that net. One option is to suppose that the reasoner could endorse the set of conclusions supported by any

[2] The credulous extensions defined here correspond, not to the *expansions* of [Touretzky, 1986], but to the *grounded expansion* (Definition 2.9, page 44). Although the concept of groundedness no longer plays any explicit role in the definition, it turns out that our extensions are grounded in the sense defined there.

one of the net's several extensions. According to this option, the reasoner will have arrived at suitable state of mind based on the information in Γ_1, for example, if he concludes that Nixon is a pacifist, or if he concludes that Nixon is not a pacifist; he cannot draw both of these conclusions, but he must draw one or the other. This option seems to be appropriate for situations in which the value of drawing conclusions is high relative to the costs involved if some of those conclusions turn out not to be correct.

A more conservative option—appropriate when the cost of error rises—is to suppose that an ideal reasoner's conclusion set based on a net with multiple credulous extensions should be determined, somehow, by the intersection of these extensions. There are two natural ways of developing this idea. We might suppose, first, that the reasoner should endorse an argument just in case it is contained in each of the network's extensions, and that he should then endorse the set of statements supported by the arguments he endorses. Or second, we might suppose that the reasoner should endorse a statement just in case it is itself supported by each of the network's extensions. As it turns out, these two approaches yield different results, since a statement may be supported in each extension, but only by different arguments. The network Γ_3 (Figure 3), for example, allows two credulous extensions, one containing the path $a \rightarrow q \rightarrow p \rightarrow t$, and the other containing the path $a \rightarrow u \rightarrow v \rightarrow t$. Each of these extensions supports the conclusion $a \rightarrow t$; so according to the second approach, the reasoner should endorse this statement. However, there is no argument in the intersection of these two extensions that supports $a \rightarrow t$; so the reasoner should not endorse this statement according to the first approach. This ambiguity in the idea of intersecting multiple credulous extensions was first pointed out in [Stein, 1989; Stein, 1990], and in [Makinson and Schlechta, 1991], where the statements supported in different extensions by different arguments are nicely described as "floating conclusions." Stein provides a polynomial-time algorithm for computing the set of statements supported in each credulous extension (although she relies on a notion of credulous extension slightly different from that given here).

Before moving on to define alternative extension concepts, we consider some fundamental results characterizing the credulous extensions.

Let us define the *generalized paths* as link sequences like paths, except that they can contain negative links anywhere, and perhaps more than one. Formally, each link is a generalized path, and if τ is a generalized path, then so are $\tau \rightarrow p$ and $\tau \not\rightarrow p$. We will say that a net is *acyclic* just in case it contains no generalized path whose initial node is identical with its end node. Now, it is natural to think of a network as "coherent" just in case it has an extension. Our first result, due to [Touretzky, 1986], provides sufficient conditions for network coherence.

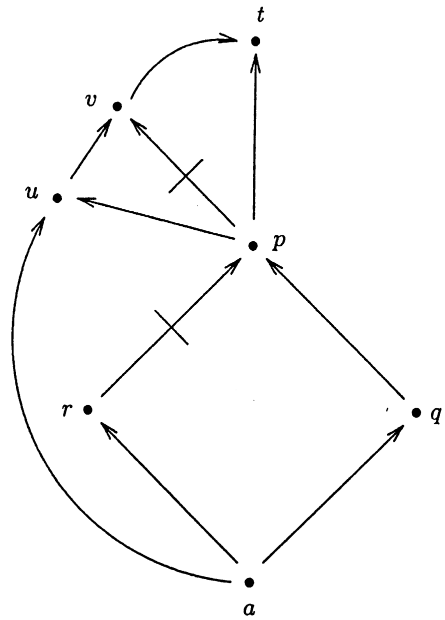

Fig. 3. Γ_3

Theorem 2.2.2. *Every acyclic network possesses a credulous extension.*

This result will follow here as a corollary to Theorem 2.2.8, presented in Section 2.2.3.

It was left as an open question in [Touretzky, 1986] whether *all* networks possess credulous extensions; but we can see now, at least for the current variant of the credulous theory, that the answer to this question is No. The cyclic net Γ_4 (Figure 4) provides a counterexample. Suppose this net had an extension—say, Φ. Of course, Φ either would or would not contain the path $(*)$ $p \rightarrow q \rightarrow r$. If Φ did contain $(*)$, it would also have to contain the path $(**)$ $p \rightarrow q \rightarrow r \rightarrow s \rightarrow t \rightarrow q$; so $(*)$ would be preempted in the context $\langle \Gamma_4, \Phi \rangle$, and therefore not inheritable. On the other hand, if Φ did not contain $(*)$, it could not contain $(**)$ either; so $(*)$ would not be preempted, and so it would be inheritable in the context $\langle \Gamma_4, \Phi \rangle$.

The next two results are analogs of those established for skeptical extensions in [Horty *et al.*, 1990]; the proofs provided there adapt easily to the credulous case.

Theorem 2.2.3. *An extension Φ of a net Γ supports both the statements*

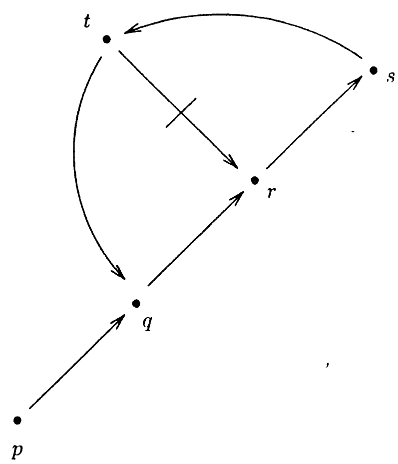

Fig. 4. Γ_4

$x \rightarrow y$ and $x \not\rightarrow y$ iff both $x \rightarrow y$ and $x \not\rightarrow y$ belong to Γ.

Theorem 2.2.4. *If Φ is an extension of a network Γ and Φ supports an atomic statement $\pi(a,p)$ $[\overline{\pi}(a,p)]$, then there is an extension Φ' of $\Gamma \cup \{\pi(a,p)\}$ $[\Gamma \cup \{\overline{\pi}(a,p)\}]$ such that Φ' supports A iff Φ supports A, for any statement A.*

The first of these (in the interesting direction) is a form of *soundness*, guaranteeing that an extension of a network will not generate any new conflicts; it will support only those conflicting statements already present in the original network. The second result assures the property of *atomic stability*—it isolates a sense in which the semantic properties of a network are unaffected when that network is supplemented with its own conclusions. This stability property is closely related to the property of cumulative monotony originally explored in [Gabbay, 1985] and [Makinson, 1989], and more recently in [Kraus *et al.*, 1990] (where stability is equivalent to a combination of the rules described as Cut and Cautious Monotonicity). Although the property was not defined formally for inheritance networks until [Horty *et al.*, 1990], it was actually the failure of atomic stability in shortest-path inheritance reasoners that provided much of the original motivation behind [Touretzky, 1986]; see, for example, the discussion of level-skips in Section 1.8 of that work.

The matter of stability is examined in greater detail in Section 5.3 of [Horty *et al.*, 1990]. An example is provided to show that *generic* stability fails for the inheritance reasoner defined there (supplementing a network with its own generic statements as conclusions may indeed affect the semantics of the net), and the same example shows that generic stability fails also for the present notion of credulous inheritance. It is suggested that it may be reasonable to require an inheritance system to exhibit atomic but not generic stability. However, the matter is still unresolved, and this middle position has been challenged from both sides: [Geffner and Verma, 1989] advocates a theory of inheritance that does not exhibit even atomic stability, while [Boutilier, 1989a] criticizes theories of the sort described in the present chapter on the grounds that they do not exhibit generic stability as well.

2.2.2 Skeptical extensions

The skeptical extensions, first defined in [Horty *et al.*, 1990], are somewhat more complicated than the credulous extensions to motivate. While the credulous extensions can be characterized directly through a fixed point equation, the skeptical extensions must be defined instead through an iterative process of moving, in a rational way, from one epistemic context to another. Beginning with a net Γ, and then at each succeeding stage in the reasoning process, we suppose that the agent repeatedly supplements his current epistemic context with the set of arguments which he is justified in accepting, but which he has not yet explicitly endorsed. This process continues until it reaches a limit—the skeptical extension of Γ—at which point the agent has explicitly endorsed all the arguments he is justified in accepting, and he can be seen also as justified in accepting all the arguments he has explicitly endorsed.

In order to flesh out this picture, we need to understand the conditions under which an agent can be said to be justified in accepting an argument. According to the theory of [Horty *et al.*, 1990], there are two such conditions.

The first represents a global constraint on the overall reasoning process: at any given stage, an agent can be justified in accepting an argument σ only if he has previously evaluated, or is currently evaluating, every other argument τ that might possibly have a bearing on the acceptability of σ (either conflicting with σ, or else figuring as a situator or preemptor in the preemption of σ). This constraint is captured by appeal to the notion of degree. Using the concept of a generalized path set out in Section 2.2.1, we can define the *degree* of a path σ in a net Γ—written, $\deg_\Gamma(\sigma)$—to be 1 if σ is a link, and otherwise, if σ is a compound path, to be the length of the longest generalized path in Γ from the initial node of σ to its end node. (Of course, this definition makes sense only for acyclic networks; the definition of skeptical inheritance applies only to these.) It is argued in [Horty *et al.*, 1990], and it should be clear anyway, that a path τ can have a bearing on the acceptability of σ only if $\deg_\Gamma(\tau) \leq \deg_\Gamma(\sigma)$. Therefore, we can enforce our global constraint on the reasoning process simply by requiring that the agent consider, and then either accept or reject, argument paths in the order of their degree: at the n-th stage of the reasoning process, the agent can consider only those arguments of degree n.

The second condition that an argument must satisfy in order to count as justified at some stage is local; the condition can be framed in terms of the epistemic context reached at that stage alone, without reference to the place of this context in the overall process of reasoning. Basically, we want to require that a skeptical reasoner can accept an argument in a context if that argument is both forcible for the reasoner, and it does not conflict with any other argument that is also forcible. We will say that an argument path meeting this condition is *permitted*, and use the symbol $\succ$ to stand for the relation of permission between a context and an argument.

Definition 2.2.5 (Permission).

Case I: σ is a direct link. Then $\langle\Gamma, \Phi\rangle \succ \sigma$ iff $\sigma \in \Gamma$.

Case II: σ is a compound path. Then $\langle\Gamma, \Phi\rangle \succ \sigma$ iff

1. $\langle\Gamma, \Phi\rangle \mid\!\sim \sigma$,
2. there is no path τ such that $\langle\Gamma, \Phi\rangle \mid\!\sim \tau$ and τ conflicts (in Γ) with σ.

Now that we have described the two conditions that a path must meet in order to count as justified, it is a straightforward matter to formalize our notion of the skeptical extension of a network as the result of repeatedly supplementing the network with justified arguments—the permitted arguments of appropriate degree.

Definition 2.2.6. Where the sequence of path sets $\Phi_1, \Phi_2, \Phi_3, \ldots$ is given by

$$\begin{array}{lcll} \Phi_1 & = & \Gamma, & \\ \Phi_{n+1} & = & \Phi_n & \cup \quad \{\sigma \; : \; deg_\Gamma(\sigma) = n+1 \; and \; \langle\Gamma, \Phi_n\rangle \succ \sigma \,\}, \end{array}$$

the path set $\bigcup_{n=1}^{\infty} \Phi_n$ is the *skeptical extension* of Γ.

Although it is based on the same motivating intuitions, the notion defined here of a skeptical extension does not quite agree with that of [Horty *et al.*, 1990]; the differences will be discussed in Section 4.3.

It should be clear that any acyclic net has a skeptical extension, that this skeptical extension is unique, and that the iterative process through which it is defined quickly reaches a limit (if k is the largest degree of any path in Γ, then Φ_k is the skeptical extension of Γ). Of course, the

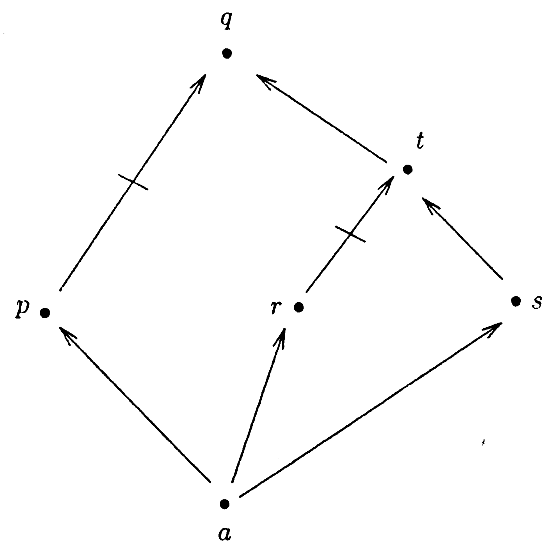

Fig. 5. Γ_5

analogs of Theorems 2.2.3 and 2.2.4—soundness and atomic stability—hold for skeptical extensions as well as credulous extensions. In addition, it is easy to see (and it will follow as a corollary to Theorem 2.2.9) that $\Phi = \{\sigma : \langle\Gamma, \Phi\rangle \mathrel{|\!\!\sim} \sigma\}$ whenever Φ is a skeptical extension of the net Γ. Just as the credulous extensions are fixed points, or equlibrium states, determined by the inheritability relation, the skeptical extensions are fixed points determined by the relation of permission.[3]

Now, what is the relation between the skeptical extension of a net and its credulous extensions? If a net possesses only one credulous extension, this will coincide with its skeptical extension as defined here. However, if a net possesses more than one credulous extension, its skeptical extension may not coincide with the intersection of these credulous extensions. This situation can occur in the case of nested diamonds, such as Γ_5 (Figure 5). Although the skeptical extension of this net contains the path $a \rightarrow p \not\rightarrow q$, the net allows a credulous extension that does not contain this path, but contains $a \rightarrow s \rightarrow t \rightarrow q$ instead.

The example of Γ_5 is drawn from [Horty *et al.*, 1990]. To the authors

[3]Note, however, that the skeptical extensions cannot be *defined* as the fixed points of the permission relation, since not all fixed points of the permission relation are skeptical extensions. The matter will be treated in the following Section 2.2.3, and also later in Section 4.3.

of that paper, this kind of example suggested that the skeptical theory presented there offered a genuinely new and alternative notion of extension, not one readily definable in terms of concepts already familiar from the credulous theory. Oddly, however, the authors of [Touretzky *et al.*, 1987] saw the matter in a somewhat different light: the gap between the skeptical extension and the intersection of credulous extensions led these authors to suspect that there might be a conceptual problem with this approach to skepticism.

It is easy to understand the reasons for this suspicion. In the case of Γ_5, for example, it is tempting to suppose, since the argument $a \rightarrow s \rightarrow t \rightarrow q$ appears in some credulous extension of the net, that this argument should be considered explicitly as one that *might* be correct; and therefore, that a truly skeptical reasoner should recognize this fact by refusing to endorse the conflicting argument $a \rightarrow p \not\rightarrow q$. This line of thought leads eventually to the view that the intersection of credulous extensions should provide the ideal for skeptical reasoning—a view that has been suggested or explicitly endorsed by a number of people, including [Boutilier, 1989a], [Geffner and Verma, 1989], [Makinson and Schlechta, 1991], and [Stein, 1989; Stein, 1990; Stein, forthcoming].

What makes it seem so natural that the intersection of credulous extensions should be thought to provide the ideal for skeptical reasoning is a particular interpretation of these extensions, as well as a particular interpretation of the relation between skepticism and credulity. The interpretation—which is most explicit in Stein's work—is that the credulous extensions of a network are to represent the possible states of the world consistent with that network, and that the goal of skeptical reasoning is to draw only those conclusions guaranteed to hold no matter which of these possible states is actual. Given this way of viewing things, of course, it does follow at once that the goal of any skeptical reasoner should be to arrive at the intersection of credulous extensions.

But this interpretation of extensions, as possible states of the world, is not the only way of understanding them. The credulous extensions could just as easily be taken to represent the ideal or equlibrium mental states of an agent for whom a credulous reasoning strategy is appropriate, an agent for whom the value of drawing conclusions suggested by a network outweighs the cost of error.[4] It is likewise possible to interpret the skeptical extensions simply as equilibrium states of an agent for whom the tradeoff between drawing conclusions and avoiding error is less heavily weighted in favor of conclusions. And given this way of understanding the extensions, as states of mind rather than states of the world, there no longer seems to be any inevitable connection between the skeptical extension of a net and

[4] This interpretation of the credulous extensions is due to Doyle [Doyle, 1985; Doyle, 1988]; a related interpretation is developed in [Horty, forthcoming].

the intersection of its credulous extensions. Why should a reasoner pursuing some skeptical strategy reject an argument just because some credulous reasoner, pursuing an entirely different strategy, accepts a conflicting argument?

In some ways, it seems best to regard these two approaches to skeptical reasoning—the intersection of extensions approach advocated by Stein and others, and the alternative approach described in the present section—simply as different and equally coherent ways of developing a shared underlying conception of skepticism.

The central feature of any skeptical reasoning strategy—what makes it *skeptical*—is the idea that an agent should refrain from accepting an argument in the face of a plausible counterargument; but this underlying idea can lead to different particular theories depending upon the manner in which the "plausible" counterarguments to a given argument are identified. One way of developing the idea is to imagine that any argument falling in a credulous extension (any argument that could be endorsed by some credulous reasoner) should count as plausible, so that a skeptical reasoner cannot endorse an argument if it conflicts with one of these. This is the option that leads to the intersection of extensions view; it is, of course, coherent and intuitively appealing, although it does force the agent, at times, to regard as plausible certain counterarguments that he himself has no reason to accept.[5] Such arguments are characterized by [Makinson and Schlechta, 1991] as "zombies"—paths that are themselves dead, though still able to kill others. It seems equally coherent, however, to develop the basic skeptical idea in a way that avoids zombies, by identifying as plausible only those arguments that actually have some force for the reasoner, in his current epistemic context; and these are the inheritable paths. The approach to skepticism described in [Horty *et al.*, 1990], and in the present section, is based on this point of view.

2.2.3 Flexible extensions

So far, we have considered only extensions that treat each of the conflicts engendered by a net uniformly, in either a credulous or a skeptical manner (regardless of the precise way in which skepticism is achieved). We have supposed that a reasoner might adopt different reasoning strategies in different situations—employing a credulous strategy when it is most important to arrive at definite conclusions, and some form of skeptical strategy when it is more important to avoid error. However, the theories considered so far do not allow the reasoner to adopt, in a single situation, a credulous

[5] This is to be understood in the sense described earlier, according to which an agent has a reason to accept an argument in a context only if that argument is constructible; in this sense, an agent in the context $\langle \Gamma_5, \Gamma_5 \rangle$ has no reason to accept the path $a \rightarrow s \rightarrow t \rightarrow q$.

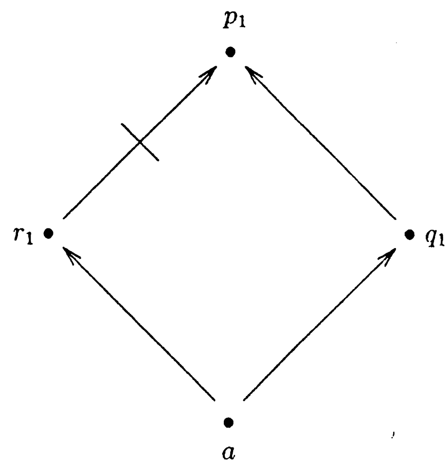

Fig. 6. Γ_6

attitude toward one conflict and a skeptical attitude toward another.

The insistence on this kind of uniformity in any given situation seems in some cases to rule out extensions containing the appropriate set of conclusions. For example, consider the two diamonds Γ_6 (Figure 6) and Γ_7 (Figure 7); and suppose that the nature of the information represented by Γ_6 makes it appropriate to adopt a credulous strategy in drawing conclusions from this net, but that the nature of the information in Γ_7 makes a skeptical strategy appropriate there. Now imagine that the reasoner is given as his initial data the net $\Gamma_6 \cup \Gamma_7$, rather than the individual nets Γ_6 and Γ_7 in isolation. What conclusions should he draw? If he adopts a credulous strategy in order to accommodate the information from Γ_6, he will be forced, inappropriately, either either to conclude that a is a p_2 or to conclude that a is not a p_2. But if he adopts a skeptical strategy in order to handle the information from Γ_7 properly, he will not be able, as he should, either to conclude that a is a p_1 or to conclude that a is not a p_1.

As long as we insist that an ideal reasoner should arrive at some extension, and extensions exhibit a uniform treatment of conflicts, then it seems that we cannot allow the reasoner to arrive at the proper set of conclusions in situations such as these, where a more flexible treatment of conflicts is called for. This problem suggests that a new extension concept—the concept of a *flexible* extension—should be introduced to characterize the appropriate conclusions in situations like this, just as the credulous and skeptical extensions characterize the appropriate conclusions in situations

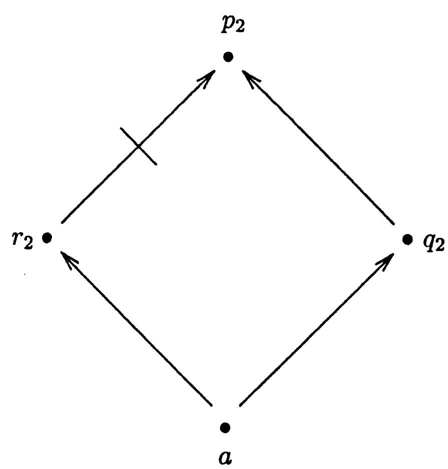

Fig. 7. Γ_7

requiring uniformly credulous or skeptical reasoning. Extensions of this kind can be defined quite simply, by taking fixed points of the permission relation.

Definition 2.2.7. The path set Φ is a *flexible extension* of the net Γ iff

$$\Phi = \{\sigma : \langle \Gamma, \Phi \rangle \mathrel{|\!\!\sim} \sigma \}.$$

Both of the intuitively appropriate sets $\Gamma_6 \cup \Gamma_7 \cup \{a \rightarrow q_1 \rightarrow p_1\}$ and $\Gamma_6 \cup \Gamma_7 \cup \{a \rightarrow r_1 \not\rightarrow p_1\}$ then turn out to be flexible extensions of the $\Gamma_6 \cup \Gamma_7$.[6]

Although the flexible extensions include path sets that are neither credulous nor skeptical extensions, it turns out that they generalize the previous extension concepts: each credulous or skeptical extension is also a flexible

[6] This simple example may give the impression that it is possible to arrive at each flexible extension of any net, as it is in this case, by partitioning the net into skeptical and credulous regions, and then joining the skeptical extension of the first region with some credulous extension of the other. However, the impression is misleading. For instance, the net Γ_5 described earlier allows the path set $\Gamma_5 \cup \{a \rightarrow s \rightarrow t\}$ as a flexible extension, resulting from a credulous approach toward the conflict at t, and then skeptical approach toward the conflict at q. But this flexible extension cannot be arrived at by partitioning the net into credulous and skeptical regions; the extension depends on the path $a \rightarrow s \rightarrow t \rightarrow q$, which moves through a conflict toward which a credulous attitude is adopted in order to reach a conflict toward which a skeptical attitude is adopted.

extension. In fact, the flexible extensions can be thought of as those fixed points arrived at by following a range of reasoning strategies, with the skeptical and credulous extensions falling at the extremes of this range.

This can be seen most easily by focusing on the iterative constructions naturally associated with each of the three extension concepts. Suppose, just as in the previous Section 2.2.2, that an agent is constructing an extension of an acyclic network Γ by considering, and then either accepting or rejecting, the potential arguments in the net stage by stage, in order of degree; however, this time, suppose that the target extension can be of any sort—credulous, skeptical, or flexible. Just as before, we will assume that in reasoning his way to the extension the agent moves through a monotone sequence $\Phi_1, \Phi_2, \Phi_3 \ldots$ of approximations, where the extension is the limit of this sequence. As before, we take $\Phi_1 = \Gamma$. At each successive stage of the reasoning process, we will assume that the agent supplements the set of arguments he has currently endorsed with some subset of the arguments of appropriate degree inheritable in his current epistemic context. Let us define

$$I_n(\Gamma, \Phi) = \{\sigma : \langle \Gamma, \Phi \rangle \mid\!\sim \sigma \text{ and } \deg_\Gamma(\sigma) = n\},$$

so that, at the stage of augmenting Φ_n to form Φ_{n+1}, the argument set from which the agent chooses is given by $I_{n+1}(\Gamma, \Phi_n)$. Of course, the agent cannot simply endorse all the arguments contained in this set, since in general, it may contain conflicts. If the agent's initial information is given by Γ_1 (the Nixon diamond), for example, then the argument set from which he will have to choose at the second stage of the reasoning process is $I_2(\Gamma_1, \Gamma_1) = \{a \to q \to p, a \to r \not\to p\}$.

In considering how, at the $(n+1)$-th stage of his reasoning process, the agent might choose from among the paths in $I_{n+1}(\Gamma, \Phi_n)$, it is useful to separate out explicitly the conflicting paths belonging to this set. Therefore, let us define $C_\Gamma(\Phi) = \{\sigma : \sigma \in \Phi \text{ and } \sigma \text{ is conflicted in } \langle \Gamma, \Phi \rangle\}$ and $S_\Gamma(\Phi) = \Phi - C_\Gamma(\Phi)$, so that, intuitively, $C_\Gamma[I_{n+1}(\Gamma, \Phi_n)]$ represents the set of conflicts among inheritable paths that must be resolved by the agent, while $S_\Gamma[I_{n+1}(\Gamma, \Phi_n)]$ represents the set of inheritable paths that are unconflicted, or safe. In addition, we will say that a subset Ξ of Ψ is *statement uniform* just in case, whenever Ξ contains any argument from Ψ supporting a particular statement, it must contain every argument from Ψ supporting that statement.

Using these ideas, the different iterative rules for moving from Φ_n to Φ_{n+1} appropriate to each of the three extension concepts can be cast in the same general form; each requires that

$$(*) \qquad \Phi_{n+1} = \Phi_n \cup S_\Gamma[I_{n+1}(\Gamma, \Phi_n)] \cup \Psi,$$

with Ψ specified as some conflict free and statement uniform subset of $C_\Gamma[I_{n+1}(\Gamma, \Phi_n)]$. Where these three versions of the rule $(*)$ differ, of course, is in the further restrictions they place on Ψ. According to the skeptical theory, Ψ is required to be the empty set; according to the credulous theory, Ψ is required to be some *maximal* conflict free subset of $C_\Gamma[I_{n+1}(\Gamma, \Phi_n)]$. The flexible theory, lying between these two extremes, places no further restrictions on Ψ—allowing it to be the empty set, a maximal conflict free subset of $C_\Gamma[I_{n+1}(\Gamma, \Phi_n)]$, or any other subset that is both conflict free and statement uniform.[7]

Let us say that $\Phi_1, \Phi_2, \Phi_3 \ldots$ is a *reasoning sequence* based on Γ if $\Phi_1 = \Gamma$ and each Φ_{n+1} is formed from Φ_n through some version of the iterative rule $(*)$; and depending on the particular version of the iterative rule adopted, let us say that the reasoning sequence is *credulous*, *skeptical*, or *flexible*. A moment's thought shows that the skeptical extension of any net coincides with the limit of the skeptical reasoning sequence based on that net, since the skeptical reasoning sequence defined here will contain exactly the same path sets as the sequence presented in Definition 2.2.6, through which the skeptical extension of a net is defined. It is less obvious, but true nonetheless, that the credulous and flexible reasoning sequences correspond in the same way to the credulous and flexible extensions.

Theorem 2.2.8. *Φ is a credulous extension of an acyclic net Γ iff Φ is the limit of a credulous reasoning sequence based on Γ.*

Theorem 2.2.9. *Φ is a flexible extension of an acyclic net Γ iff Φ is the limit of a flexible reasoning sequence based on Γ.*

In addition to establishing the existence of credulous and flexible extensions for acyclic nets, these results show that the iterative characterization of those extensions coincides with the fixed point characterization.[8]

---

[7] Of course, the empty set and any maximal conflict free subset of $C_\Gamma[I_{n+1}(\Gamma, \Phi_n)]$ are automatically statement uniform, but the requirement does need to be imposed separately in the case of flexible reasoning. To see the need for this, take $\Gamma = \Gamma_1 \cup \{a \rightarrow s, s \rightarrow p\}$, so that $I_2(\Gamma, \Gamma) = \{a \rightarrow q \rightarrow p, a \rightarrow s \rightarrow p, a \rightarrow r \not\rightarrow p\}$. Without the requirement of statement uniformity, there would be nothing to prevent the agent from selecting at this stage the argument $a \rightarrow q \rightarrow p$, but not the argument $a \rightarrow s \rightarrow p$ as well. However, the set $\Gamma \cup \{a \rightarrow q \rightarrow p\}$ is not a flexible extension of Γ.

[8] Theorem 2.2.8 should be compared to Theorem 2.11 of [Touretzky, 1986], which shows in the same way (through the construction of a reasoning sequence) that any net whose IS-A subgraph is acyclic has a credulous extension. Touretzky's theorem has broader scope, since nets that are cyclic by the present standards might still be IS-A acyclic. On the other hand, Touretzky's theorem shows only that some extension can be reached as the limit of a reasoning sequence; and not all of the credulous extensions can be approximated through reasoning sequences of the sort he defines. The present result shows that the credulous extensions coincide with the limits of reasoning sequences of the sort defined here.

3 Theories of mixed inheritance

The theories developed in Section 2 apply only to purely defeasible networks, a severe expressive limitation, as pointed out in [Brachman, 1985]. The present section shows how these theories can be extended to apply also to networks containing both strict and defeasible links mixed together. In a sense, the real point of this section is that no new ideas are involved in extending the theories of defeasible inheritance to this broader range of networks; it is simply a matter of generalizing the previous ideas to the new environment. Still, the generalization is complicated, and it is worth working through it in detail.

This section follows an expository track similar to that of Section 2, first setting out a notion of inheritability appropriate for mixed networks, and then using this new notion to define the extensions. The notion of inheritability presented here for mixed nets is itself a mixture, combining the treatment of defeasible inheritability from Section 2.1 with the theory of strict inheritance developed in [Thomason *et al.*, 1986]. We begin, then, by reviewing briefly this approach to strict inheritance.

3.1 Strict inheritability

According to this theory, strict inheritance is really very simple: each strict network Γ has a unique extension Φ, which contains exactly those paths that can be constructed from the links in Γ, and so a unique theory, containing exactly those statements supported by Φ. The unique extension of the network Γ_8 (Figure 8), for example, contains the paths $a \Rightarrow s \Rightarrow r$ and $p \Rightarrow q \not\Leftrightarrow r \Leftarrow s$. Suppose we interpret the nodes in this net so that $p =$ starlings, $q =$ birds, $r =$ mammals, $s =$ dogs, and $a =$ Rover. Then the first of these paths show that the unique theory of Γ_8 contains the conclusion that Rover is a mammal ($a \Rightarrow r$); the second shows that it contains the conclusion that no starlings are dogs ($p \not\Leftrightarrow s$).

It is important to note that this analysis of strict inheritance, although straightforward, is not the standard view. Strict networks contain only strict links, each of which can be represented, as explained in Section 1.3, by a formula of classical logic. It may seem natural, then, to use classical logic itself to provide a semantics for such a network—by identifying the network with the set of formulas that translate its links, and then defining a statement as supported by the network just in case it belongs to the deductive closure of that set. This idea, which seems to be the standard view, is due originally to [Hayes, 1979]. To see that it differs from the analysis proposed here, consider, for example, the net Γ_9 (Figure 9). This net would be translated into the set $\{Pa, \neg Pa, \neg Qa\}$. Since the set is inconsistent, any statement at all belongs to its classical deductive closure;

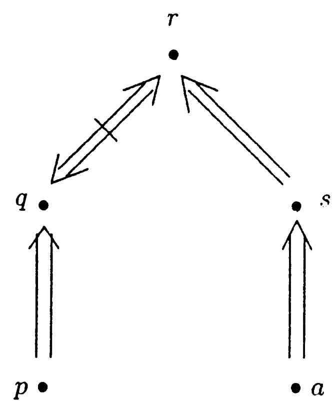

Fig. 8. Γ_8

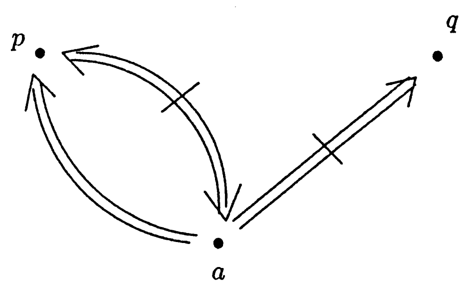

Fig. 9. Γ_9

so according to the standard view, the theory of Γ_9 should be taken to contain every statement—including, say, Qa. According to the analysis of [Thomason *et al.*, 1986], however, the theory of Γ_9 does not contain Qa; the extension of this net contains no positive path from a to q, and in fact provides uncontested evidence that $\neg Qa$.

It is, in some ways, a delicate matter to decide between the present analysis and the traditional analysis of [Hayes, 1979]. One is always free to regard a strict network simply as a notational variant of some classical theory, so that the traditional analysis would be appropriate. Still, there seems to be some value in taking seriously the graph-based nature of inheritance reasoners, which derive conclusions corresponding only to actual paths. The problem then is to see how we can make logical sense of such a reasoner—by designing an appropriate logic, rather than forcing the reasoner to conform to the standards of an already-existing logic. This task is carried out for strict networks in [Thomason *et al.*, 1986], which provides both a proof theory for path-based inheritance reasoning and an interpretation of the resulting logic in a four-valued model based on that of Belnap [Belnap, 1977a; Belnap, 1977b].

The proof theory is a calculus in the style of [Gentzen, 1934] for proving sequents of the form $\Gamma \vdash A$, where Γ is a set of statements (a net) and A a statement (a link). Informally, such a sequent is supposed to mean that A is derivable from Γ. The sequent calculus contains as its only *structural rule* the schema

$$A \vdash A,$$

where A is a literal—that is, a link of the form $a \Rightarrow p$ or $a \nLeftrightarrow p$. This gives us our axioms. In addition, we have the following *logical rules*, for introducing both $\Rightarrow$ and $\nLeftrightarrow$, on the right and on the left of the turnstile.

$$\frac{\Gamma^a, a \Rightarrow p \vdash a \Rightarrow q}{\Gamma^a \vdash p \Rightarrow q} \vdash\Rightarrow$$

$$\frac{\Gamma \vdash a \Rightarrow p \quad \Delta, a \Rightarrow q \vdash A}{\Gamma, \Delta, p \Rightarrow q \vdash A} \Rightarrow\vdash \qquad \frac{\Gamma \vdash a \nLeftrightarrow q \quad \Delta, a \nLeftrightarrow p \vdash A}{\Gamma, \Delta, p \Rightarrow q \vdash A} \Rightarrow\vdash'$$

$$\frac{\Gamma^a, a \Rightarrow p \vdash a \nLeftrightarrow q}{\Gamma^a \vdash p \nLeftrightarrow q} \vdash \nLeftrightarrow$$

$$\frac{\Gamma \vdash a \Rightarrow p \quad \Delta, a \nLeftrightarrow q \vdash A}{\Gamma, \Delta, p \nLeftrightarrow q \vdash A} \nLeftrightarrow \vdash \qquad \frac{\Gamma \vdash a \Rightarrow q \quad \Delta, a \nLeftrightarrow p \vdash A}{\Gamma, \Delta, p \nLeftrightarrow q \vdash A} \nLeftrightarrow \vdash'$$

In the rules $\vdash\Rightarrow$ and $\vdash \nLeftrightarrow$, Γ^a is supposed to represent a collection of formulas not containing a. We do need both the rules $\Rightarrow\vdash$ and $\Rightarrow\vdash'$ to capture the meaning of $\Rightarrow$ on the left of the turnstile; neither will do alone. Likewise, both $\nLeftrightarrow \vdash$ and $\nLeftrightarrow \vdash'$ are necessary.

We provide here a sample proof, of the sequent $p \Rightarrow q, q \nLeftrightarrow r \vdash p \nLeftrightarrow r$, simply in order to illustrate these rules.

$$\frac{\dfrac{\dfrac{a \Rightarrow p \vdash a \Rightarrow p \quad a \Rightarrow q \vdash a \Rightarrow q}{a \Rightarrow p, p \Rightarrow q \vdash a \Rightarrow q} \Rightarrow\vdash \qquad a \nLeftrightarrow r \vdash a \nLeftrightarrow r}{p \Rightarrow q, q \nLeftrightarrow r, a \Rightarrow p \vdash a \nLeftrightarrow r} \nLeftrightarrow \vdash}{p \Rightarrow q, q \nLeftrightarrow r \vdash p \nLeftrightarrow r} \vdash \nLeftrightarrow$$

The techniques of [Thomason *et al.*, 1986] can be used to establish the following theorem, which shows that this sequent calculus is both sound and complete with respect to the analysis presented above of strict inheritance.

Theorem 3.1.1. *The sequent $\Gamma \vdash A$ is provable iff A belongs to the theory of Γ.*

The interpretation of this logic relies on the set $\mathcal{T} = \{\{\mathrm{T}\}, \{\mathrm{F}\}, \emptyset, \{\mathrm{T}, \mathrm{F}\}\}$ of truth values. Following Belnap, we use these values to represent four possible states of a knowledge base with respect to a proposition: the state of possessing evidence for the proposition and no evidence to the contrary; the state of possessing evidence against the proposition and no evidence to the contrary; the state of possessing no evidence either for or against the proposition; the state of possessing evidence both for the proposition and against it well. This explanation should suggest why it is natural to take the power set of $\{\mathrm{T}, \mathrm{F}\}$ as the set of truth values: if $X \in \mathcal{T}$ is the truth value for some proposition, then $\mathrm{T} \in X$ just in case there is evidence for the proposition, and $\mathrm{F} \in X$ just in case there is evidence against it.

A *valuation* v on the language of strict links can be defined as follows. Relative to a domain D, the valuation assigns an individual $v(a)$ in D to each individual term a of the language, and a function $v(p)$ from D to $\mathcal{T}$ to each generic term p. Where v is a valuation, v^{d}/a is the valuation like v for all terms other than a, but which assigns the value d to a. The following rules extend v to the entire language.

- $v(a \Rightarrow p) = [v(p)](v(a))$.
- $v(a \not\Leftrightarrow p) = \mathrm{Not}(v(pa))$, where $\mathrm{Not}(\{\mathrm{T}\}) = \{\mathrm{F}\}$, $\mathrm{Not}(\{\mathrm{F}\}) = \{\mathrm{T}\}$, $\mathrm{Not}(\emptyset) = \emptyset$, and $\mathrm{Not}(\{\mathrm{T}, \mathrm{F}\}) = \{\mathrm{T}, \mathrm{F}\}$.
- $v(p \Rightarrow q) = \{\mathrm{T}\}$ if for all $\mathrm{d} \in \mathrm{D}$, we have $\mathrm{T} \in v^{\mathrm{d}}/a(a \Rightarrow q)$ if $\mathrm{T} \in v^{\mathrm{d}}/a(a \Rightarrow p)$ and $\mathrm{F} \in v^{\mathrm{d}}/a(a \Rightarrow p)$ if $\mathrm{F} \in v^{\mathrm{d}}/a(a \Rightarrow q)$; and $v(p \Rightarrow q) = \emptyset$ otherwise.
- $v(p \not\Leftrightarrow q) = \{\mathrm{T}\}$ if for all $\mathrm{d} \in \mathrm{D}$, we have $\mathrm{F} \in v^{\mathrm{d}}/a(a \Rightarrow q)$ if $\mathrm{T} \in v^{\mathrm{d}}/a(a \Rightarrow p)$ and $F \in v^{\mathrm{d}}/a(a \Rightarrow p)$ if $\mathrm{T} \in v^{\mathrm{d}}/a(a \Rightarrow q)$; and $v(p \not\Leftrightarrow q) = \emptyset$ otherwise.

Given this interpretation, the notion of semantic implication is defined in the usual way.

Definition 3.1.2. Γ *semantically implies* A iff, for all valuations v, if $\mathrm{T} \in v(B)$ for all $B \in \Gamma$, then $\mathrm{T} \in v(A)$.

It can then be shown that this kind of four-valued implication characterizes the sequent calculus.

Theorem 3.1.3. *The sequent $\Gamma \vdash A$ is provable iff Γ semantically implies A.*

Together with Theorem 3.1.1, or course, this means that the four-valued semantic implication also characterizes the path-based notion of strict inheritance described above.

3.2 Mixed inheritability

We turn now to the task of defining a new notion of inheritability for mixed networks, by combining the theory of strict inheritability just reviewed with the treatment of defeasible inheritability from Section 2.1.

It will be useful to focus first on the special case of mixed inheritability for paths ending in defeasible links; this is, after all the most interesting case, since it is here that we isolate the conditions under which we can draw an inference using defeasible information. For paths of this kind, the account of mixed inheritability is similar in its overall structure to the account of defeasible inheritability presented earlier: a simple path (or link) will be classified as inheritable if it is contained in the net; a compound path will be classified as inheritable if it is constructible, but neither conflicted nor preempted. In fact, the notion of constructibility for compound paths can simply be carried over from the defeasible case: Definition 2.1.1 applies without change. However, the ideas of conflict and preemption must be modified in the mixed environment to accommodate the presence of strict links.

In defeasible networks, all conflicts share a simple form: they involve paths with identical initial nodes, identical end nodes, and opposite polarity. But the presence of strict links introduces the possibility of less direct conflicts, even among purely defeasible paths. As an illustration, consider Γ_{10} (Figure 10). Here it seems reasonable, in light of the strict segment $r \Rightarrow s \Rightarrow t$, to regard $p \rightarrow q \rightarrow r$ and $p \rightarrow u \rightarrow v \not\rightarrow t$ themselves as conflicting paths, even though they do not share an end node. Imagine, for example, that $r =$ dogs, $s =$ mammals, and $t =$ animals, so that the strict segment tells us that *all* dogs are animals. In the context of Γ_{10}, then, the path $p \rightarrow q \rightarrow r$, which represents an argument to the effect that p's are dogs, carries with equal force the conclusion that p's are animals; so it conflicts with $p \rightarrow u \rightarrow v \not\rightarrow t$, which represents an argument that p's are not animals.

What this example shows is that two paths can represent conflicting arguments, even if they have different end nodes, when one of the paths clashes with a strict consequence of the other. Of course, such strict consequences can themselves be classified as positive or negative. Let us define

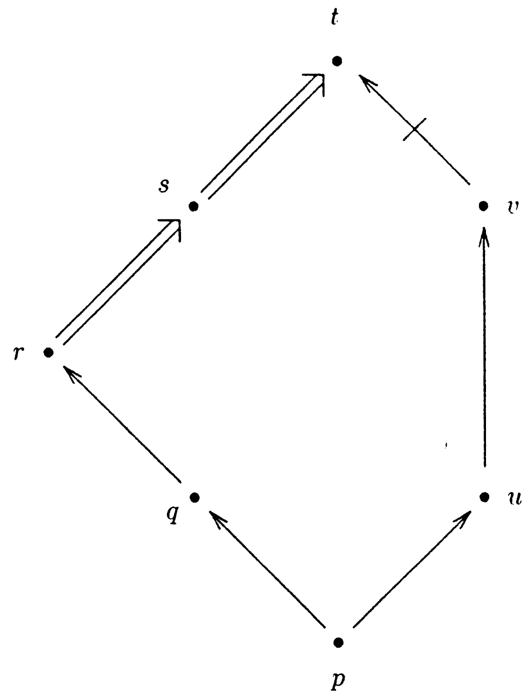

Fig. 10. Γ_{10}

$$\kappa_\Gamma(x) = \{x\} \cup \{y : \Gamma \text{ allows a strict positive path from } x \text{ to } y\},$$
$$\overline{\kappa}_\Gamma(x) = \{y : \Gamma \text{ allows a strict negative path from } x \text{ to } y\},$$

so that $\kappa_\Gamma(x)$ and $\overline{\kappa}_\Gamma(x)$ represent the positive and negative strict consequences attributed to x by Γ—the set of properties that x must possess, according to Γ, and the set of properties that x cannot possess. It is then natural to extend our conception of conflicting paths as follows.

Definition 3.2.1 (Mixed conflict). A path $\pi(x, \sigma, y)$ *conflicts in* Γ with any path of the form $\overline{\pi}(x, \tau, m)$ for $m \in \kappa_\Gamma(y)$, and also with any path of the form $\pi(x, \tau, m)$ for $m \in \overline{\kappa}_\Gamma(y)$.

Given this generalized notion of conflict, the characterization of conflicted paths from Definition 2.1.3 can carry over unchanged.

As it turns out, even this revised formulation is not adequate to capture the intuitive notion of conflict in mixed nets containing distinct but strictly equivalent nodes. This can be seen by considering Γ_{11} (Figure 11). According to the definition just presented, the paths $x \rightarrow u \rightarrow y$ and

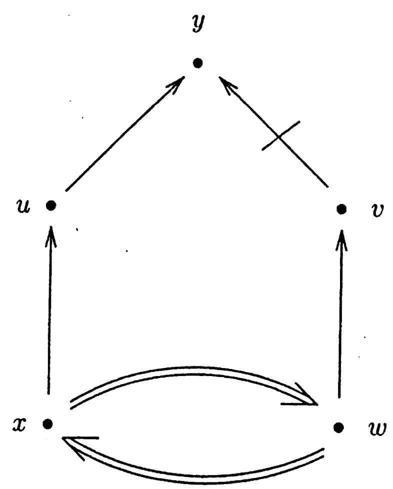

Fig. 11. Γ_{11}

$w \rightarrow v \not\rightarrow y$ do not conflict with each other in this net, since the node x is distinct from w. But from an intuitive point of view, it does seem that these should count as conflicting paths, and that a consistent extension should not contain both, since the net tells us that x's are strictly equivalent to w's.

There are several ways to handle this problem. The first is simply to disallow nets containing strict cycles of the kind contained in Γ_{11}, just as we earlier disallowed nets containing defeasible cycles. This solves the problem, of course, by ruling out the offending cases; but the solution is extreme, since it is sometimes useful to be able to represent in a network distinct but strictly equivalent concepts. (For example, the designer of a knowledge base containing some mathematical information might wish to distinguish the concepts of triangular and trilateral plane figures, while recording that an instance of each is necessarily an instance of the other.) A better solution would be to reformulate the notion of mixed conflict so that it gives the intuitively correct results in nets containing strictly equivalent nodes; but this strategy also is problematic. The presence of such nodes affects the treatment of preemption as well as conflict in mixed nets. Both notions would have to be adjusted to handle the attendant difficulties; the adjustments are complicated, and in the case of preemption, they lead to a definition that is nearly unintelligible. Moreover, the inference algorithms naturally suggested by the definitions designed to deal directly with the problems of strictly equivalent nodes are more cumbersome than those based on definitions that are able to ignore these problems.

A third solution, which combines some advantages of the others, is to treat networks containing strictly equivalent nodes by proxy, providing a semantics for these networks, without actually complicating the definitions of conflict and preemption to deal with equivalent nodes, by deriving their meaning from others without strict equivalences. This can be done quite simply by associating with each net Γ possibly containing strictly equivalent nodes another net Γ^* in which each set of strictly equivalent nodes from Γ has been collapsed into a single node. First, we let

$$[x]_\Gamma = \{y : y \in \kappa_\Gamma(x) \text{ and } x \in \kappa_\Gamma(y)\},$$

so that $[x]_\Gamma$ represents the set of nodes strictly equivalent to x in the net Γ. These strict equivalence classes from Γ form the individual nodes of Γ^*, and the links between these nodes are induced in the natural way: we take $\pi([x]_\Gamma, [y]_\Gamma) \in \Gamma^*$ iff there exist $u \in [x]_\Gamma$ and $v \in [y]_\Gamma$ such that $\pi(u, v) \in \Gamma$, and likewise $\overline{\pi}([x]_\Gamma, [y]_\Gamma) \in \Gamma^*$ iff there exist $u \in [x]_\Gamma$ and $v \in [y]_\Gamma$ such that $\overline{\pi}(u, v) \in \Gamma$. The net Γ^*, of course, is guaranteed to be free from strictly equivalent nodes, and so we can use the simpler definitions of conflict and preemption to characterizing its extensions; but because of its construction, it seems that this net can be taken also to provide a semantics for Γ. The easiest way to connect the two is by extending slightly the relation of support between argument paths and statements, so that paths belonging to extensions of Γ^* can be said to support statements based on Γ: where $u \in [x]_\Gamma$ and $v \in [y]_\Gamma$, we will say that a positive path $\pi([x]_\Gamma, \sigma, [y]_\Gamma)$ supports the corresponding statement $\pi(u, v)$, and also that a negative path $\overline{\pi}([x]_\Gamma, \sigma, [y]_\Gamma)$ supports the statement $\overline{\pi}(u, v)$. In this way, the extensions of Γ^* can be used to determine the theories of Γ.

Throughout the remainder of this section, we will avoid any complications arising from nets containing distinct but strictly equivalent nodes, supposing that these nets can be handled by proxy in the way suggested. In implementational terms, the move from a net Γ to its proxy Γ^* can be thought of as a kind to preprocessing to be carried out before querying takes place. It may be expensive, but the expense may be recoverable during query time by the use of simpler algorithms.

We turn now to the matter of preemption. Just as the presence of strict links allows for the possibility of new kinds of conflicts in mixed nets, it provides also for the possibility of new relations of preemption. To see this, imagine that the agent is supplied with the net Γ_{12} (Figure 12) as his initial information, where a = Hermann, p = persons born in America, q = native speakers of German, r = persons born in Pennsylvania, and s = native speakers of Pennsylvania Dutch. Under this interpretation, Γ_{12} contains the information that Hermann is a particular speaker of Pennsylvania Dutch, that every speaker of Pennsylvania Dutch speaks German (since Pennsylvania Dutch is a dialect of German), that German speakers

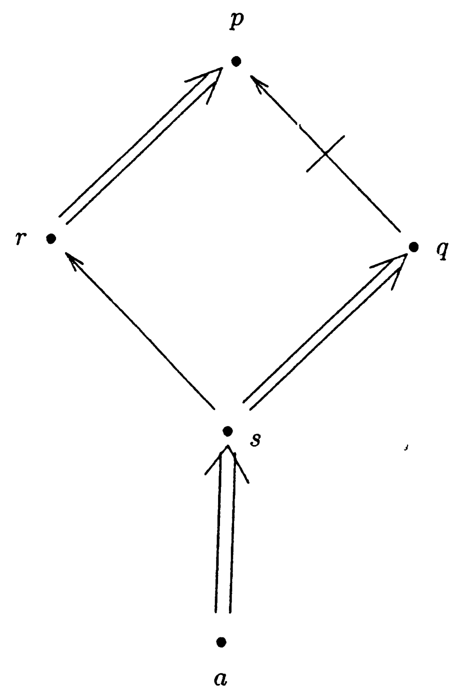

Fig. 12. Γ_{12}

tend not to be born in America, that speakers of Pennsylvania Dutch tend to be born in Pennsylvania, and that everyone born in Pennsylvania is born in America.

Now suppose that the agent has already drawn the conclusion that Hermann is a native speaker of German, reasoning his way to the epistemic context $\langle \Gamma_{12}, \Phi \rangle$, where $\Phi = \Gamma_{12} \cup \{a \Rightarrow s \Rightarrow q\}$. In this context, the paths $a \Rightarrow s \rightarrow r$ and $a \Rightarrow s \Rightarrow q \not\rightarrow p$ are both constructible. Neither is conflicted, yet they stand in conflict with each other, because $p \in \kappa_{\Gamma_{12}}(r)$. Of course, the agent should not be free in this case to endorse whichever one of these argument paths he chooses. Intuitively, it looks as if the path $a \Rightarrow s \Rightarrow q \not\rightarrow p$, representing the argument that Hermann was not born in America since he is a native speaker of German, should be preempted in the context—since the fact that his dialect is Pennsylvania Dutch seems to provide a better argument to the contrary. Without modification, however, our previous analysis of preemption does not give us this result. A path can be preempted only if there is information to the contrary that is both more specific and immediate; and, although s does provide more specific information than q, the path $s \rightarrow r \Rightarrow p$ does not represent *immediate* information to the contrary—at least, not according to our earlier standards, which hold that immediate information can be carried only by individual links.

Evidently, it is this last requirement concerning the nature of immediate information that needs to be modified. In a purely defeasible environment, it makes good sense to say that immediate information can be carried only by single links: a compound path represents only a tentative argument, which can itself be undermined. In the environment of mixed nets, however, certain kinds of compound defeasible paths can legitimately be thought to carry immediate information—namely, those paths consisting of a single defeasible link followed by a strict end segment, of any length. In Γ_{12}, for example, the path $s \rightarrow r \Rightarrow p$ should be thought of as telling us immediately that speakers of Pennsylvania Dutch are born in America. Even by our earlier standards, $s \rightarrow r$ counts as an immediate statement of the fact that speakers of Pennsylvania Dutch are born in Pennsylvania, and the strict extension $r \Rightarrow p$ simply tells us that *everyone* born in Pennsylvania is born in America.

Generalizing from this example leads to the following notion of preemption for mixed nets.

Definition 3.2.2 (Mixed preemption). A positive path of the form $\pi(x, \sigma, u) \rightarrow y$ is *preempted* in the context $\langle \Gamma, \Phi \rangle$ iff there exist nodes v, m such that (i) either $v = x$ or there is a path of the form $\pi(x, \tau_1, v, \tau_2, u) \in \Phi$, and (ii) either (a) $v \not\rightarrow m \in \Gamma$ and $m \in \kappa_\Gamma(y)$ or (b) $v \rightarrow m \in \Gamma$ and $m \in \overline{\kappa}_\Gamma(y)$. A negative path of the form $\pi(x, \sigma, u) \not\rightarrow y$ is *preempted* in the context $\langle \Gamma, \Phi \rangle$ iff there exist nodes v, m such that (i) either $v = x$ or there is a path of the form $\pi(x, \tau_1, v, \tau_2, u) \in \Phi$, and (ii) $v \rightarrow m \in \Gamma$ and $y \in \kappa_\Gamma(m)$.

Again, when a positive path of the form $\pi(x, \sigma, u) \rightarrow y$ is preempted in accord with this definition, we refer to v as the preempting node, and we describe the path $\pi(x, \tau_1, v, \tau_2, u)$ as the situator. The preemptor is either the path $\pi(x, \tau_1, v) \not\rightarrow m$ or the path $\pi(x, \tau_1, v) \rightarrow m$, depending on whether the preemption relation is realized through clause (ii.a) or (ii.b). This vocabulary applies likewise to the preemption of negative paths.

At this point, we are able to define the concept of mixed inheritability for paths ending in defeasible links: happily, the previous Definition 2.1.5 will itself do the job, once the defeasible notions of conflict and preemption appealed to in this definition are replaced with their mixed variants. In order to extend this concept of mixed inheritability to the entire range of mixed paths, we first introduce some notation for analyzing these paths according to their structure.

Any path σ from a mixed network can be divided into the subpaths $\mathrm{Str}(\sigma)$ and $\mathrm{Def}(\sigma)$, where $\mathrm{Str}(\sigma)$ is the maximal strict end segment of σ, and $\mathrm{Def}(\sigma)$ is the defeasible initial segment that results from truncating

Str(σ) from σ. (Example: if σ is $x \Rightarrow y \rightarrow p \not\Leftrightarrow r \Leftarrow s$, then Str($\sigma$) is $p \not\Leftrightarrow r \Leftarrow s$ and Def(σ) is $x \Rightarrow y \rightarrow p$.) Using this notation, we can exhaustively classify the mixed paths in the following way. Such a path σ might consist of a non-null defeasible initial segment followed by a non-null strict end segment, in which case we would have $\sigma \neq \mathrm{Def}(\sigma)$ and $\sigma \neq \mathrm{Str}(\sigma)$. Alternatively, the path might be entirely strict, in which case we would have $\sigma = \mathrm{Str}(\sigma)$, or it might have no strict end segment at all, in which case we would have $\sigma = \mathrm{Def}(\sigma)$. The definition of mixed inheritability treats paths in accord with this case structure.

Definition 3.2.3 (Mixed inheritability).

Case A: $\sigma \neq \mathrm{Def}(\sigma)$ and $\sigma \neq \mathrm{Str}(\sigma)$. Then $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$ iff $\mathrm{Str}(\sigma) \in \Phi$ and $\mathrm{Def}(\sigma) \in \Phi$.

Case B: $\sigma = \mathrm{Str}(\sigma)$. Then $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$ iff σ is a path constructed from links in Γ.

Case C-I: $\sigma = \mathrm{Def}(\sigma)$ and σ is a direct link. Then $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$ iff $\sigma \in \Gamma$.

Case C-II: $\sigma = \mathrm{Def}(\sigma)$ and σ is a compound path. Then $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$ iff

1. σ is constructible in $\langle \Gamma, \Phi \rangle$,
2. σ is not conflicted in $\langle \Gamma, \Phi \rangle$,
3. σ is not preempted in $\langle \Gamma, \Phi \rangle$.

Here, Case A reduces the question of inheritability for defeasible paths with strict end segments to two simpler question: inheritability for strict paths, and inheritability for defeasible paths without strict end segments. These are treated in Cases B and C.

It turns out, as it should, that this notion of mixed inheritability is a conservative generalization of the ingredient theories of strict and defeasible inheritability from in [Thomason *et al.*, 1986] and Section 2.1—in the sense that it agrees with these theories when applied to purely strict or defeasible nets. The verification of this fact is reassuringly simple. It is evident from Case B of the definition that this account agrees with the treatment of [Thomason *et al.*, 1986] when applied to strict nets. If Γ is a purely defeasible net, on the other hand, then only Cases C-I and C-II of the current definition come into play. Case C-I is identical to the corresponding clause from Definition 2.1.5. And when Γ is defeasible, we have $\overline{\kappa}_\Gamma(x) = \emptyset$ and $\kappa_\Gamma(x) = \{x\}$ for any node x. Under these conditions, the notions of mixed conflict and preemption presented in Definitions 10 and 3.2.2 turn out to be logically equivalent to the corresponding defeasible notions from Definitions 2.1.2 and 2.1.4; so Case C-II from Definition 3.2.3 is likewise identical to the corresponding clause from Definition 2.1.5.

3.3 Constructing extensions

We now use this notion of mixed inheritability to characterize the credulous, skeptical, and flexible extensions of a mixed network. The concepts leading up to credulous mixed inheritance are drawn from [Horty, 1991]; the treatment of skeptical mixed inheritance is motivated by the same intuitions as [Horty and Thomason, 1988], but differs slightly in details.

3.3.1 Fixed point definitions

Two of our three extension concepts—the credulous and flexible extensions—are defined through fixed point equations, and we turn first to these.

The credulous extensions of defeasible nets were characterized as fixed points of the defeasible inheritability relation. The credulous extensions of mixed nets can be defined in exactly the same way, using the notion of mixed inheritability: Definition 2.2.1 applies without change. The flexible extensions of defeasible nets were characterized as fixed points of the relation of defeasible permission. However, this relation of permission was defined in terms of defeasible inheritability. If we appeal instead to mixed inheritability, then Definition 2.2.5 as it stands yields a relation of mixed permission, and we can characterize the flexible extensions of mixed nets as the fixed points of this new relation: again, Definition 2.2.7 applies without change.

Since, as we have seen, mixed inheritability is a conservative generalization of defeasible inheritability, this way of defining the credulous and flexible mixed extensions results also in a conservative generalization of the definitions for the defeasible case. In addition, it is easy to develop mixed analogs to the Theorems 2.2.2 through 2.2.4 to characterize the credulous and flexible mixed extensions.

Theorem 2.2.2 tells us that each acyclic defeasible net has an extension, where the acyclic nets were defined as those without cyclic generalized paths. In the present mixed environment, we need to broaden the notion of a generalized path, and then adjust the notion of cyclicity used in this definition. Formally, we will say now that each link is a generalized path; and that if τ is a generalized path, then so are $\tau \rightarrow p$, $\tau \not\rightarrow p$, $\tau \Rightarrow p$, $\tau \not\Leftrightarrow p$, and $\tau \Leftarrow p$. (Example: both $p \not\rightarrow q \Leftarrow r \not\rightarrow s \Leftarrow t$ and $a \rightarrow p \Leftarrow q$ are generalized paths, but neither is a path.) Among the generalized paths, we classify as *defeasible* those that contain at least one defeasible link. And we say, finally, that a mixed net is *acyclic* if it is free from defeasible cycles—that is, if it allows no defeasible generalized path whose initial node is identical with its end node.

Given this new characterization of cyclicity, the statement of Theorem 2.2.2 can be carried over into the environment of mixed nets without change; in its new form, this theorem will follow as a corollary to Theorem 3.3.3. Notice that the new ideas necessary for incorporating this theorem into the mixed environment generalize the earlier versions: in a purely defeasible environment, these broader ideas classify exactly the same link sequences described earlier as generalized paths, and they classify exactly the same nets as cyclic. Therefore, the restatement of Theorem 2.2.2 using these more general ideas serves only to broaden its applicability; it does not change the original meaning of the theorem.

The point of Theorem 2.2.3 is that, in a purely defeasible environment, a net allows an extension supporting conflicting statements only if that net itself contains those conflicting statements. Again, the main idea behind this theorem applies also to the mixed environment; but its generalization results in a more complicated formulation, due to the more complicated notion of conflict in mixed nets.

Theorem 3.3.1. *An extension Φ of a net Γ supports both the statements $\pi(x, y)$ and $\overline{\pi}(x, y)$ iff (i) a link of the form $\pi(x, u)$ belongs to Γ where $y \in \kappa_\Gamma(u)$, and either (ii) a link $\pi(x, v)$ belongs to Γ where $y \in \overline{\kappa}_\Gamma(v)$ or (iii) a link $\overline{\pi}(x, v)$ belongs to Γ where $v \in \kappa_\Gamma(y)$.*

This theorem suggests also a syntactic criterion for inconsistency: a mixed net Γ can be classified as *inconsistent* just in case (i) holds, along with either (ii) or (iii). A very general treatment of inconsistency in knowledge bases containing mixed strict and defeasible information is presented in [Goldszmidt and Pearl, 1991], and the general criterion for inconsistency put forth there agrees with our simple syntactic criterion when it is restricted to the language of mixed networks (Goldszmidt and Pearl, personal communication).[9]

Finally, the statement of Theorem 2.2.4 on atomic stability, formulated earlier only for defeasible nets, carries over into the mixed environment without change.

3.3.2 Skeptical extensions

The skeptical extensions of defeasible nets were defined through the interaction between two concepts: permission and degree. The agent was imagined to step through the arguments in a net in the order of their degree, accepting at each stage only the permitted arguments. As we have

[9] Notice that our syntactic criterion for inconsistency is correct only for networks that do not contain strictly equivalent nodes. The net $\Gamma = \{x \Rightarrow w, w \Rightarrow x, x \rightarrow y, w \not\rightarrow y\}$, for example, is intuitively inconsistent, and inconsistent also according to the theory of [Goldszmidt and Pearl, 1991], but not according to our syntactic criterion. Inconsistency for networks containing strictly equivalent nodes can, however, be defined by proxy. Such a net can be classified as inconsistent just in case the network that results from when its strictly equivalent nodes have been collapsed (as explained above, in the discussion of mixed conflict) is itself inconsistent according to our syntactic criterion.

seen, the notion of permission can easily be extended to the mixed environment. In order to characterize the skeptical extensions of mixed networks, therefore, we need only develop an appropriate concept of mixed degree.

In the purely defeasible environment, the degree of a path was defined as the length of the longest generalized path from its initial node to its end node. Adapting to the present environment, we first recall from the previous Section 3.3.1 that the notion of a generalized path has itself been broadened, and second, we define the *defeasible length* of these new generalized paths as follows: if a generalized path does not contain a strict initial segment, then its defeasible length is simply the number of defeasible links in the path; if a generalized path does contain a strict initial segment, then its defeasible length is the number of defeasible links in the path augmented by 1. (Example: the generalized path $r \rightarrow s \Rightarrow t \rightarrow u$ has a defeasible length of 2, since it contains two defeasible links and no strict initial segment; the generalized path $p \Rightarrow q \Rightarrow r \rightarrow s \Rightarrow t \rightarrow u$ is 3, since it contains a strict initial segment along with two defeasible links.) Using these ideas, we can now define the *defeasible degree* of a path σ in a mixed net Γ to be 1 if σ is either a link or a strict path, and otherwise, if σ is a compound defeasible path, to be the greatest defeasible length of any generalized path in Γ from the initial node of σ to its end node. (Example: the defeasible degree of $p \rightarrow q \rightarrow r$ in the net Γ_{10} is 3 since the generalized path from p to r in Γ_{10} whose defeasible length is greatest is $p \rightarrow u \rightarrow v \not\rightarrow t \Leftarrow s \Leftarrow r$, with a defeasible length of 3.) Again, we must limit our consideration to acyclic mixed networks, so that the notion of defeasible degree makes sense.

This notion of defeasible degree is a straightforward generalization of the idea of degree from Section 2.2.2. However, it is not yet quite appropriate as a standard for specifying the order in which a skeptical reasoner should consider the arguments from a mixed net; in the present environment, the appropriate standard must carry just a bit more information. Basically, what we need to know of an argument path, in addition its defeasible degree, is whether or not that path possesses a strict end segment. Therefore, we define the *mixed degree* of a path σ in a net Γ—written $\text{mdeg}_\Gamma(\sigma)$—as a pair $\langle n, v\rangle$. The first component of the pair, n, tells us the defeasible degree of σ in Γ. The second component tells us, simply, whether or not σ possesses a strict end segment: by convention, we let $v = 0$ if σ does not possess a strict end segment, and $v = 1$ if it does. (Example: $\text{mdeg}_{\Gamma_{10}}(p \rightarrow q \rightarrow r) = \langle 3, 0\rangle$, but $\text{mdeg}_{\Gamma_{10}}(p \rightarrow q \rightarrow r \Rightarrow s) = \langle 3, 1\rangle$.) We impose a lexical ordering on the mixed degrees by giving priority to the first component: $\langle n, v\rangle < \langle n', v'\rangle$ iff either $n < n'$ or $n = n'$ and $v < v'$. The idea behind this ordering is that defeasible degree is the primary standard determining the order in which argument paths are considered—but of two paths identical in defeasible degree, one with and one without a strict end segment, the path lacking the strict end segment is considered first.

Using this concept of mixed degree, together with the notion of permission appropriate for mixed nets, the skeptical extensions of these nets can be defined as follows.

Definition 3.3.2. Where the sequence of path sets $\Phi_1^0, \Phi_1^1, \Phi_2^0, \Phi_2^1, \Phi_3^0, \Phi_3^1, \ldots$ is given by

$$\begin{array}{lclcl} \Phi_1^0 & = & \Gamma & \cup & \{\sigma : \mathrm{Str}(\sigma) = \sigma \text{ and } \sigma \text{ constructed from } \Gamma\}, \\ \Phi_n^1 & = & \Phi_n^0 & \cup & \{\sigma : \mathrm{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle \text{ and } \langle \Gamma, \Phi_n^0\rangle \mathrel{|\!\sim} \sigma\}, \\ \Phi_{n+1}^0 & = & \Phi_n^1 & \cup & \{\sigma : \mathrm{mdeg}_\Gamma(\sigma) = \langle n+1, 0\rangle \text{ and } \langle \Gamma, \Phi_n^1\rangle \mathrel{|\!\sim} \sigma\}, \end{array}$$

the path set $\bigcup_{n=1}^{\infty} \Phi_n^1$ is the *skeptical extension* of Γ.

The reasoning sequence defined here is supposed to exemplify the same general principles as that of Definition 2.2.6, but of course it is more complicated. At the first stage of the reasoning process, Φ_1^0, the agent accepts the individual links contained in a net along with each entirely strict path that can be constructed from the links in Γ; these strict paths can be accepted at once, because nothing can possibly interfere with them. After that, the agent begins a process of interleaving strict with defeasible inference. At each stage Φ_n^1, he will accept all the strict extensions of the arguments he has already accepted; it turns out that each of these strict extensions will be permitted. And then at each stage Φ_{n+1}^0, he will consider some of the paths (those of appropriate degree) that extend the arguments he has already accepted by one defeasible step, and he will accept those that are permitted in the context. The definitions of inheritability, and so permission, are arranged in such a way that all the hard work occurs at this stage; these definitions are built to "look ahead," so that once the agent has accepted an argument path ending in a defeasible link, he will be able to accept any of its strict extensions without further thought.

Again, it should be clear that the characterization of mixed skeptical extensions in Definition 3.3.2 is a conservative generalization of the previous Definition 2.2.6: not only is the mixed permission relation involved a conservative generalization of the defeasible permission relation, but applied to purely defeasible nets, the present notion of mixed degree would have the reasoner stepping through the argument paths in exactly the order of their ordinary degree. Again, the sequence will actually reach its limit quickly: if $\langle n, v\rangle$ is the largest mixed degree of any path in Γ, then Φ_n^v will be the skeptical extension of Γ. And again, the appropriate mixed analogs of Theorems 2.2.2 through 2.2.4 will characterize these extensions.

3.3.3 Iteration and fixed points

The credulous and flexible extensions of mixed nets were given a fixed point characterization in Section 3.3.1. Here we show also how to characterize these extensions iteratively, as limits of the appropriate reasoning sequences. We follow the train of thought set out at the end of Section 2.2.3.

As in the treatment of skeptical mixed inheritance, we suppose that the agent moves through a monotone sequence $\Phi_1^0, \Phi_1^1, \Phi_2^0, \Phi_2^1, \Phi_3^0, \Phi_3^1, \ldots$ of approximations to an extension; but here we adjust the sequence so that it can approximate credulous or flexible extensions as well. As before, we assume that at the first stage of the reasoning process, the agent accepts each individual link and each strict path contained in the net: $\Phi_1^0 = \Gamma \cup \{\sigma : \mathrm{Str}(\sigma) = \sigma$ and σ constructed from $\Gamma\}$. Now let us introduce the operator I_n^v, where

$$I_n^v(\Gamma, \Phi) = \{\sigma \ : \ \langle \Gamma, \Phi \rangle \mathrel{|\!\!\sim} \sigma \text{ and } \mathrm{mdeg}(\sigma) = \langle n, v \rangle \ \},$$

to aid in the definition of the successive stages. Again, these successive stages interleave strict with defeasible reasoning. At the strict stages, the agent simply adopts all strict extensions of the arguments he has already accepted:

$$(*) \qquad \Phi_n^1 = \Phi_n^0 \cup I_n^1(\Gamma, \Phi_n^0).$$

The defeasible stages are more complicated, since here the agent may have to decide among conflicting defeasible arguments. The decision is subject to different constraints depending on whether he is following a credulous, skeptical, or flexible reasoning sequence; but again, the different versions of the iterative rule can be cast in the the same general form. Using the operators C_Γ and S_Γ defined in Section 2.2.3, each version requires that

$$(**) \qquad \Phi_{n+1}^0 = \Phi_n^1 \cup S_\Gamma[I_{n+1}^0(\Gamma, \Phi_n^1)] \cup \Psi,$$

where Ψ is specified as some conflict free subset of $C_\Gamma[I_{n+1}^0(\Gamma, \Phi_n^1)]$. The credulous theory requires Ψ to be a maximal conflict free subset of $C_\Gamma[I_{n+1}^0 (\Gamma, \Phi_n^1)]$; the skeptical theory requires Ψ to be the empty set. As before, the flexible theory allows Ψ to be any subset of $C_\Gamma[I_{n+1}^0(\Gamma, \Phi_n^1)]$ at all that is both conflict free and statement uniform, but the concept of statement uniformity needs to be generalized to handle the complications introduced by mixed nets. We will call a subset Ξ of Ψ *statement uniform* in the mixed net Γ if it satisfies the following condition: for $\sigma \in \Psi$, whenever each path in Ψ that conflicts in Γ with σ is itself conflicted in Ξ, then $\sigma \in \Xi$.

Following the pattern of Section 2.2.3, we will say that a sequence $\Phi_1^0, \Phi_1^1, \Phi_2^0, \Phi_2^1, \Phi_3^0, \Phi_3^1, \ldots$ is a *mixed* reasoning sequence based on Γ if (i) $\Phi_1^0 = \Gamma \cup \{\sigma : \mathrm{Str}(\sigma) = \sigma$ and σ is constructed from $\Gamma\}$, (ii) each set Φ_n^1 is formed from Φ_n^0 through the iterative rule $(*)$, and (iii) each set Φ_{n+1}^0 is formed from Φ_n^1 through some version of the iterative rule $(**)$. The reasoning sequence is *credulous*, *skeptical*, or *flexible* depending on the particular version of $(**)$ that is adopted. Again, it should be clear that the skeptical reasoning sequences defined here are identical to the sequences from Definition 3.3.2, through which the skeptical extensions are defined. In addition, the limit points of the credulous and flexible reasoning sequences coincide with the these extensions.

Theorem 3.3.3. *Φ is a credulous extension of Γ iff Φ is the limit of a credulous mixed reasoning sequence based on Γ.*

Theorem 3.3.4. *Φ is a flexible extension of Γ iff Φ is the limit of a flexible mixed reasoning sequence based on Γ.*

It follows as a corollary to these theorems that credulous and flexible extensions exists for acyclic mixed networks.

4 Discussion

This section explores some alternatives in the design of path-based inheritance theories, as well as some more general issues in nonmonotonic reasoning suggested by these theories and their alternatives. We consider the proper direction of argument, the nature of preemption, yet another characterization of skeptical reasoning, and some plausible constraints on translational interpretations. (Except for a brief mention of strict information in the discussion of preemption, all of this discussion focuses on purely defeasible nets; most of the problems presented by the study of inheritance reasoning can be seen more clearly in this environment.) Finally, we consider a couple of additional concerns relevant to the application of this work in knowledge representation.

4.1 Decoupling and stability

The definitions of inheritability considered so far are based on a *forward chaining* notion of constructibility, according to which inheritable paths are constructed from the bottom up. This approach to path construction runs counter to the picture of inheritance networks as structures in which properties are thought of as flowing downward in the graph, from more general kinds to more specific kinds, and then finally to individuals. On the other hand, the forward chaining notion of construction is appropriate when one wants to emphasize the analogy between paths and arguments—since arguments, at least as they are usually represented (say, by proof sequences), tend to move from the beginning forward.

The forward chaining approach to path construction was first adopted in [Horty *et al.*, 1990]. Earlier, the approach had been explicitly considered and rejected in [Touretzky, 1986], on the grounds that it allowed a phenomenon described there as "decoupling," which seemed to present difficulties. Let us say that a positive path $\pi(x, y, \sigma, z)$ is *decoupled* in a path set Φ if $\pi(y, \sigma, z) \notin \Phi$ (and likewise for negative paths). It is easiest to understand the difficulties involved in allowing extensions to contain decoupled paths through a particular example, such as Γ_{13} (Figure 13), from [Touretzky, 1986]. Given a notion of inheritability based on a forward chaining construction—that is, according to the definitions of Section 2—this net allows as a credulous extension the path set $\Gamma_{13} \cup \{n \rightarrow r \not\rightarrow p, a \rightarrow n \rightarrow q \rightarrow p\}$. The extension is problematic, however, since it seems to suggests that a is a p in virtue of being an n, while insisting that n's themselves are not p's; the conclusions we draw about a in this extension are not properly coupled with the conclusions we draw about n.

Another problem resulting from the forward chaining treatment is illustrated in Γ_{14} (Figure 14), which allows, according to the definitions of Section 2, a credulous extension containing both the path $a \rightarrow n \rightarrow r \not\rightarrow p$ and the path $b \rightarrow n \rightarrow q \rightarrow p$. What is odd about this extension is that it supports different conclusions about a and b—telling us that a is a p while b is not—even though the original net gives us exactly the same information about these two individuals. Again, it is precisely the possibility of decoupling that is responsible for this oddity: if our conclusions about a and b were both properly coupled to our conclusions about n, they would also have to agree with each other.

Since the kind of problems illustrated by these two nets involve choices among conflicting paths, they will not arise on the skeptical approach to inheritance described in Section 2.2.2. Likewise, the problems do not seem to be terribly serious for the approach to credulous reasoning that takes the intersection of credulous extensions as the real object of concern, since the worst of the peculiarities tend to be filtered out in the passage from individual credulous extensions to their intersections. The problems do appear to be serious, however, for the approach to credulous reasoning that admits the individual credulous extensions themselves as legitimate states of mind for an ideal reasoner. For this reason, Touretzky decided to rule out the possibility of decoupling by basing his definition of inheritability on the following *double chaining* treatment of constructible paths, rather than the forward chaining construction of Section 2.

Definition 4.1.1 (Constructibility: alternative definition). A compound positive path $\pi(x_1, \ldots, x_n)$ is constructible in the context $\langle\Gamma, \Phi\rangle$ iff $\pi(x_1, \ldots, x_{n-1}) \in \Phi$ and $\pi(x_2, \ldots, x_n) \in \Phi$. A compound negative path $\overline{\pi}(x_1, \ldots, x_n)$ is constructible in the context $\langle\Gamma, \Phi\rangle$ iff $\pi(x_1, \ldots, x_{n-1}) \in \Phi$ and $\overline{\pi}(x_2, \ldots, x_n) \in \Phi$.

It should be clear that the result of replacing the forward chaining notion of constructibility with the alternative double chaining notion in the definition of defeasible inheritability (Definition 2.1.5) is enough to enforce coupling. No extension could then contain a path decoupled in that extension; and in particular, the problematic extensions described above for Γ_{13} and Γ_{14}

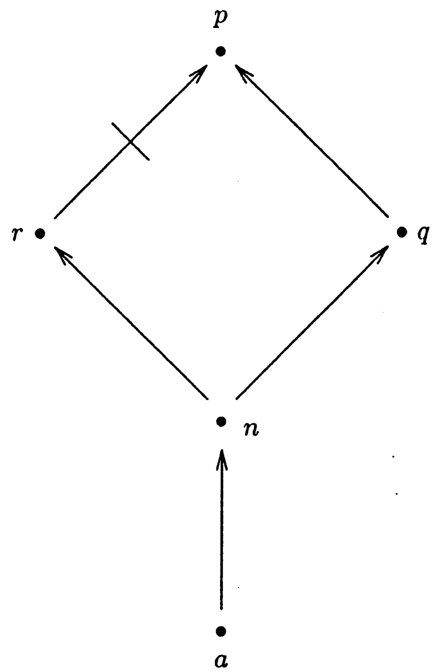

Fig. 13. Γ_{13}

could no longer arise.

Unfortunately, the insistence on coupling, and the use of a double chaining notion of constructibility to ensure coupling, themselves introduce further difficulties, which can be illustrated by the net Γ_{15} (Figure 15). In fact, if our definition of inheritability is based on the double chaining notion of path construction, this net presents problems for both the skeptical and the credulous approaches to inheritance.

On the skeptical approach, the unique extension of this net contains neither $n \rightarrow r \not\rightarrow p$ nor $n \rightarrow q \rightarrow p$; in the face of conflicting evidence, the skeptical reasoner does not conclude either that n's are p's, or that they are not. However, the individual a is a particular n for which the reasons for believing that n's are not p's is explicitly canceled. It seems intuitively correct, therefore, that the reason for believing that n's are p's should apply undisturbed in the case of a, and that the skeptical extension of this net should contain the path $a \rightarrow n \rightarrow q \rightarrow p$. The forward chaining notion of constructibility allows such an extension, but since the path $a \rightarrow n \rightarrow q \rightarrow p$ is decoupled in the extension, the double chaining notion rules it out.

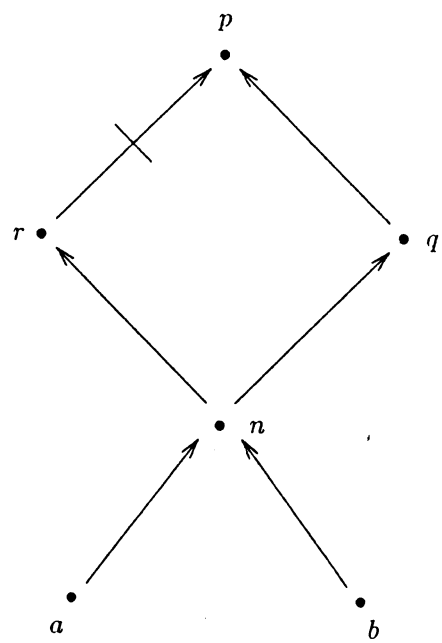

Fig. 14. Γ_{14}

On the credulous approach to inheritance, the problem presented by Γ_{15} is of another sort: this net shows that adopting the double chaining notion of path construction violates the condition of atomic stability described in Theorem 2.2.4. Given double chaining, the net allows the set $\Phi = \Gamma_{15} \cup \{n \rightarrow r \not\rightarrow p, a \rightarrow n \rightarrow q\}$ as a credulous extension. This extension, of course, supports the statement $a \rightarrow q$; but since any credulous extension of the net $\Gamma_{15} \cup \{a \rightarrow q\}$ will contain the path $a \rightarrow q \rightarrow p$, this new net will not allow a credulous extension supporting exactly the same set of statements as Φ. This violation of atomic stability seems especially serious once one recalls, as pointed out in Section 2.2.1 of the present chapter, that much of the original motivation for the credulous theory of [Touretzky, 1986] was to avoid the kind of instability associated with shortest-path inheritance reasoners.

Apparently, the treatment of coupling presents a choice between unattractive options. If we try to ensure coupling through the double chaining notion of path construction, we sacrifice desirable conclusions in the case of skeptical reasoning, and in the case of credulous reasoning, we violate the condition of atomic stability. If we allow decoupling by adopting the for-

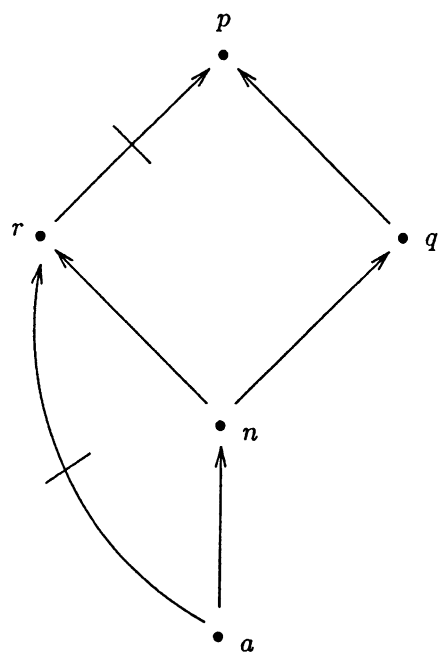

Fig. 15. Γ_{15}

ward chaining notion of constructibility, we are forced also to accept certain peculiar extensions, such as the extensions described above for Γ_{13} and Γ_{14}. It may be possible, as Makinson (personal communication) has suggested, to thread our way between these unattractive options—allowing decoupling in situations such as Γ_{15}, where it seems natural and well-motivated, but not in situations such as Γ_{13} and Γ_{14}, where the decoupling seems unmotivated or gratuitous. However, a workable proposal has not yet been developed.

4.2 Varieties of preemption

One of the most important features of an inheritance system is its treatment of preemption—the idea that a reason for accepting a defeasible argument should be overridden in any context offering a conflicting reason that is both immediate and, in some sense, better or more specific. As we have seen in Section 3.2, the notion of an immediate reason from defeasible networks needs to be modified in the presence of strict information, but these modifications are uncontroversial. Unfortunately, the general framework of path-based inheritance also allows several options for interpreting the conditions under which one reason can be thought of as better than another; and it is not really clear which is correct, or whether different ones may be appropriate for different reasoning tasks.

This section lists a few of the most prominent possibilities; we examine them in a simplified setting, putting aside the complications involved in generalizing the notion of an immediate reason to accommodate strict links. Let us suppose that an agent is in the context $\langle\Gamma,\Phi\rangle$, where Φ contains a path $\pi(x,\sigma,u)$ and Γ contains a link $u \rightarrow y$. The agent then has a reason to accept the path $\pi(x,\sigma,u) \rightarrow y$; this path is constructible. Let us suppose also, however, that Φ contains some path of the form $\pi(x,\tau,v)$, where $v \not\rightarrow y$ belongs to Γ. In this situation, the agent is faced with reasons for accepting conflicting arguments. Focusing on the positive case (the negative case is symmetric), we describe some of the different conditions under which the path $\pi(x,\sigma,u) \rightarrow y$ might be thought of as preempted by $\pi(x,\tau,v) \not\rightarrow y$. Of course, there is no question here that $v \not\rightarrow y$ provides immediate information contrary to that supplied by $u \rightarrow y$. Therefore, the different approaches to preemption result from different ways of explaining how the node v must be situated with respect to x and u in order to count as a preempting node; or alternatively, different ways of specifying the conditions under which the link $v \not\rightarrow y$ might be treated as a better reason than $u \rightarrow y$ for drawing conclusions about x.

4.2.1 Strict subsumption

One option is to suppose that $v \not\rightarrow y$ should count as a better reason than $u \rightarrow y$ just in case the class of v's is *strictly subsumed* by the class of u's. This idea leads to the following view of preemption.

- A positive path $\pi(x,\sigma,u) \rightarrow y$ is preempted in the context $\langle\Gamma,\Phi\rangle$ iff there is a node v such that (i) Φ contains both a path of the form $\pi(x,\tau_1,v)$ and a *strict* path of the form $\pi(v,\tau_2,u)$, and (ii) $v \not\rightarrow y \in \Gamma$.

This option has not actually been incorporated into any of the defeasible inheritance reasoners discussed in the literature, but it has been advocated by a number of people in conversation, and also adopted in some of the defeasible logics developed by [Nute, 1988].

The strict subsumption view of preemption has the advantage of extreme simplicity. If a network reasoner is divided into separate components for strict and defeasible reasoning, then on this view, all the specificity relations among rules or reasons can be calculated in advance by the strict component, before any defeasible inferences are drawn. Moreover, the view accords with our intuitions in many of the central and most natural examples, where preemption usually does seem to depend on a relation of strict specificity. According to the strict subsumption view, for example, preemption is ruled out in the original motivating case of Γ_2 (the Tweety Triangle).

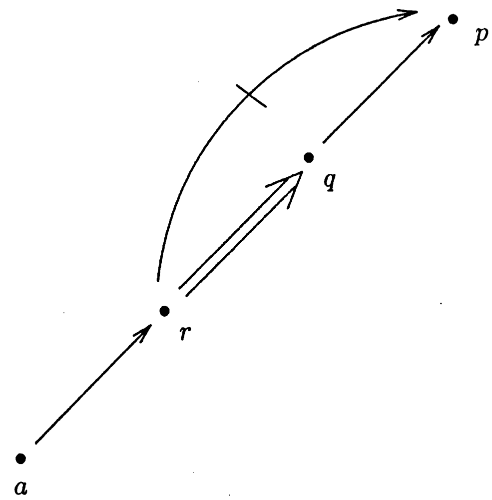

Fig. 16. Γ_{16}

But in fact, the informal interpretation of that net is represented more accurately by Γ_{16} (Figure 16), where again a = Tweety, r = penguins, q = birds, and p = flying things; and here, since the class of penguins is entirely included within the class of birds, the strict subsumption view tells us correctly that $r \not\rightarrow p$ is a better reason than $q \rightarrow p$.

Unfortunately, there are cases in which preemption does seem to be mediated by defeasible subsumption. For example, the nodes in Γ_2 can be given the following interpretation, due to [Reiter and Criscuolo, 1981]: a = Zack, r = high school dropouts, q = adults, and p = people with jobs. Since the class of dropouts is not entirely subsumed by the class of adults, the strict subsumption view tells us that $r \not\rightarrow p$ and $q \rightarrow p$ are incomparable as reasons for drawing conclusions about Zack. Depending on the treatment adopted for conflicts, this view would then lead either to multiple extensions, with Zack employed in one and unemployed in another, or else to skepticism concerning Zack's employment. However, it seems from an intuitive viewpoint that the path $a \rightarrow r \rightarrow q \rightarrow p$ should be preempted, and that the net should have only one extension, in which Zack is unemployed.

4.2.2 General subsumption

Because of examples like this, it is natural to suppose that preemption can be based on defeasible as well as strict relations of subsumption. The

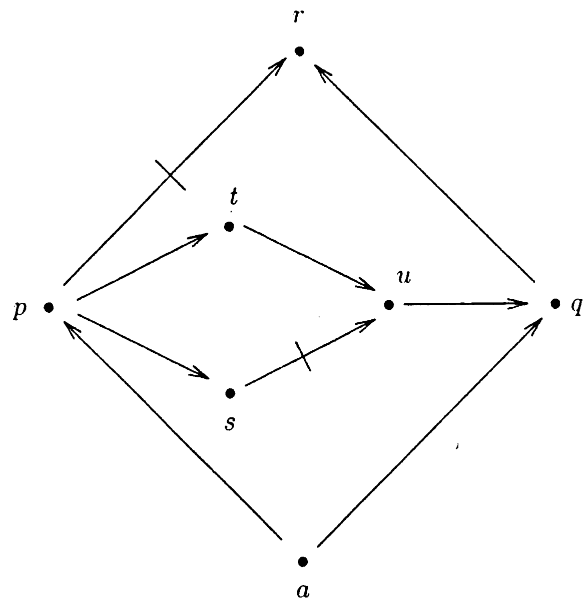

Fig. 17. Γ_{17}

easiest way of achieving this effect is by allowing the link $v \not\rightarrow y$ to count as a better reason than $v \rightarrow y$ as long as v's *tend to be* u's, even if some are not. The idea is captured formally simply by dropping from the previous proposal the requirement that the path from v to u should be strict.

- A positive path $\pi(x, \sigma, u) \rightarrow y$ is preempted in the context $\langle \Gamma, \Phi \rangle$ iff there is a node v such that (i) Φ contains both a path of the form $\pi(x, \tau_1, v)$ and a path of the form $\pi(v, \tau_2, u)$, and (ii) $v \not\rightarrow y \in \Gamma$.

This treatment seems to be the most accurate way of incorporating formally into the path-based framework the informal idea that reasons should be prioritized according to specificity, and then in the case of conflict more specific reasons should prevail.

As soon as we suppose that preemption can be mediated by defeasible relations of specificity, everything becomes more complicated, but also more interesting. It is no longer possible for the priority relations among rules to be calculated in advance by the strict component of an inheritance reasoner, and then simply fed into the defeasible component, since these priorities themselves will now depend on the results of defeasible reasoning. The system must calculate both defeasible conclusions and relations of priority among rules as it goes along, and each calculation will appeal to the results

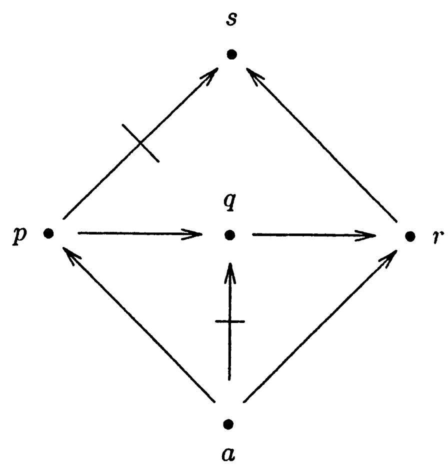

Fig. 18. Γ_{18}

of the other. Moreover, because priority relations depend on the results of defeasible reasoning, and these results can vary in theories allowing multiple extensions, it follows also that the priority relations among rules can vary from extension to extension. Consider, for example, the net Γ_{17} (Figure 17). Looking at the inner diamond, a credulous reasoner could choose to accept either the path $p \rightarrow s \not\rightarrow u$, or else the path $p \rightarrow t \rightarrow u \rightarrow q$. In those extensions in which he makes the first of these choices, the conflicting rules $p \not\rightarrow r$ and $q \rightarrow r$ will be incomparable, and so the agent is free also to choose between $a \rightarrow p \not\rightarrow r$ and $a \rightarrow q \rightarrow r$; but if he makes the second choice, the rule $p \not\rightarrow r$ is given a greater priority than $q \rightarrow r$, and so $a \rightarrow q \rightarrow r$ is preempted.

4.2.3 Off-path preemption

Although it may be the simplest, the general subsumption treatment is not the only way of basing preemption on a defeasible relation of specificity. Another option is to suppose that, in order to count $v \not\rightarrow y$ as a better reason than $u \rightarrow y$ for drawing conclusions about a x, we should require not only a path from v to u, but a path *from x through v to u*. For reasons to be discussed shortly, this idea has been described as "off-path" preemption in [Touretzky *et al.*, 1987]; it is the treatment incorporated into our official Definition 2.1.4, but we will repeat it here for convenience.

- A positive path $\pi(x, \sigma, u) \rightarrow y$ is preempted in the context $\langle \Gamma, \Phi \rangle$ iff there is a node v such that (i) Φ contains a path of the form

$\pi(x, \tau_1, v, \tau_2, u)$, and (ii) $v \not\to y \in \Gamma$.

This approach to preemption was first suggested in [Sandewall, 1986], and advocated also [Stein, 1990]; a very similar treatment was suggested independently by Cross (personal communication) to the authors of [Horty *et al.*, 1990], and adopted there.

Unlike the approach based on general subsumption, the off-path treatment of preemption does not lead to a prioritization among rules that applies to all nodes at once; instead, it leads to a prioritization that can vary as we reason about one node or another. On this treatment, then, the specificity relation that mediates preemption is not a two-place relation, telling us simply that one rule provides better information than another; it is a three-place relation, telling us only that one rule provides better information than another when we are reasoning about a particular node. In addition, the off-path approach to preemption is intensional in a way that the general subsumption treatment is not. According to the general subsumption treatment, we can determine the priority rankings among rules in a context once we know the statements (both strict and defeasible) supported in that context. According to the off-path approach, however, it is not enough to know only the statements supported in order to determine priority rankings among rules; in addition, we must know the particular arguments through which these statements are supported.

The differences between these two approaches can be illustrated by the net Γ_{18} (Figure 18), where p = lawyers, q = ambitious people, r = accomplished people, s = people who are useful to society, and a = Ann.[10] Suppose we have reached the context $\langle \Gamma_{18}, \Phi \rangle$ where $\Phi = \Gamma_{18} \cup \{p \to q \to r\}$. According to the two-place view embodied in the general subsumption definition, the rule $p \not\to s$ concerning lawyers is then supposed to represent better information in general than the rule $r \to s$ concerning accomplished people, since the context tell us through the path $p \to q \to r$ that lawyers are a defeasible subset of accomplished people. Therefore, the path $a \to r \to s$ is preempted, and we must accept instead the argument $a \to p \not\to s$, telling us that Ann is socially useless. According to the three-place view embodied in the off-path treatment, however, the rule $p \not\to s$ is not supposed to represent better information about Ann than the rule $r \to s$, since, although the net allows a path from a to p and a path from p to r, it does not allow a path from a through p to r. We know that Ann is a lawyer, that she is accomplished, and that lawyers are accomplished; but since we cannot conclude that Ann is accomplished *in virtue of* being a lawyer, the rule about lawyers does not provide better information than the rule about accomplished people in our reasoning about this particular individual.

[10] This node labeling is adapted from an earlier example of Ginsberg's (personal communication), which displayed a similar attitude toward lawyers.

4.2.4 On-path preemption

The final option we consider is the original treatment of preemption, due to [Touretzky, 1986]. According to the previous approach, a path of the form $\pi(x, \sigma, u) \rightarrow y$ is supposed to be preempted in the presence of a path $\pi(x, \tau_1, v, \tau_2, u)$ and a link $v \not\rightarrow y$. This treatment is described as off-path because there is no requirement that the preempting node v should lie on the initial segment $\pi(x, \sigma, u)$ of the path that it preempts. By contrast, Touretzky's original treatment is described in [Touretzky *et al.*, 1987] as "on-path" preemption. Strictly speaking, this is a misnomer, because the treatment does not actually require v to lie on the path $\pi(x, \sigma, u)$; however, it does require a very close relation between the preempting node and preempted path.

Suppose $\pi(x, \sigma, u)$ has the form $\pi(x_1, \ldots, x_n)$. Following Touretzky, we will say that the node v is an *intermediary* to this path in the path set Φ if $v = x_i$ for some $1 \leq i < n$, or else Φ contains a path of the form $\pi(x_1, \ldots, x_i, y_1, \ldots, y_j, x_{i+1}, \ldots, x_n)$ for some $1 \leq i < n$ such that $v = y_k$ for some $1 \leq k \leq j$. We can think of the intermediaries to a path as those nodes lying either on the path, or else "almost" on the path, in the sense that they lie on another path like the original except that it interpolates a longer argument between adjacent nodes of the original. We can illustrate the idea by contrasting Γ_{19} (Figure 19) with Γ_{20} (Figure 20). Suppose we are working against the background of a path set that contains at least $a \rightarrow p \rightarrow q$. Consider first the path $a \rightarrow q$ in Γ_{19}. The node p is an intermediary to this path, since it lies on $a \rightarrow p \rightarrow q$. Because this longer path interpolates an argument between adjacent nodes of the original, it can be thought of simply as expanding an argument that is abbreviated there. Now consider the path $a \rightarrow r \rightarrow q$ in Γ_{20}. The node p is not classified as an intermediary to this path. Since $a \rightarrow p \rightarrow q$ does not place p between adjacent nodes of the original path $a \rightarrow r \rightarrow q$, it cannot be thought of as expanding an argument that was implicit in the original; instead, it must be considered as an entirely different line of reasoning.

Using this notion of an intermediary, the on-path treatment of preemption is given as follows.

- A positive path $\pi(x, \sigma, u) \rightarrow y$ is preempted in the context $\langle\Gamma, \Phi\rangle$ iff there is a node v such that (i) v is an intermediary to the path $\pi(x, \sigma, u)$ in the path set Φ, and (ii) $v \not\rightarrow y \in \Gamma$.

According to this treatment, the path $a \rightarrow q \rightarrow s$ will be preempted in the case of Γ_{19}, since p an intermediary. But $a \rightarrow r \rightarrow q \rightarrow s$ will not be preempted in Γ_{20}, since p is not an intermediary there. This result about Γ_{20} has been challenged by [Sandewall, 1986], primarily on the basis of a node labeling which makes it seems intuitively that the path $a \rightarrow r \rightarrow q \rightarrow s$ should be preempted. However, an alternative labeling has been supplied

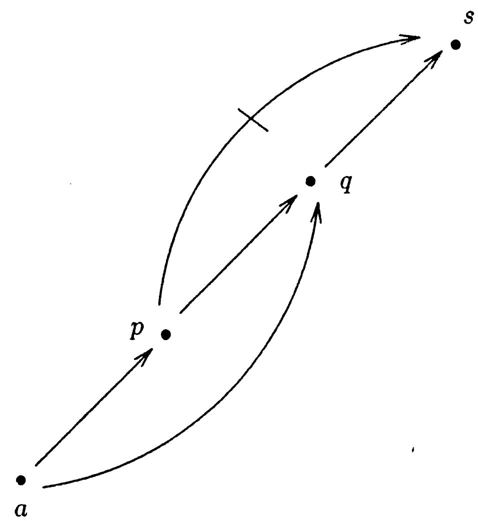

Fig. 19. Γ_{19}

in [Touretzky *et al.*, 1987] that makes preemption seem less plausible in this net; and the results of on-path reasoning have been defended also in [Boutilier, 1989a].

On-path preemption is complicated; and apart from arguments based on particular node labelings, it is sometimes hard to see any more general rationale for it. Touretzky does not really try to motivate the details of this treatment in [Touretzky, 1986]; and as we have seen, the the informal idea that "more specific" information dominates seems to support various other approaches, depending on whether specificity is interpreted as a strict relation, or as a two- or three- place defeasible relation.[11] In fact, the on-path approach treats preemption as a highly intensional notion, sensitive

[11] What Touretzky actually does say to motivate his treatment of preemption is somewhat misleading. In Section 8.2 of [Touretzky, 1986], entitled "Inferential distance in 25 words or less," he describes the idea as follows:

> The essential intuition behind inheritance exceptions is: subclasses override superclases. Briefly stated, the inferential distance rule says that A may view B as a subclass of C iff A has an inference path *via* B to C, and not vice versa.

But these two sentences actually seems to support two different views. The first (subclasses override superclases) suggests what is described here as the general subsumption treatment; the second seems to supports the off-path approach. Neither really describes the more complicated on-path treatment that is adopted in the mathematical portions of this work.

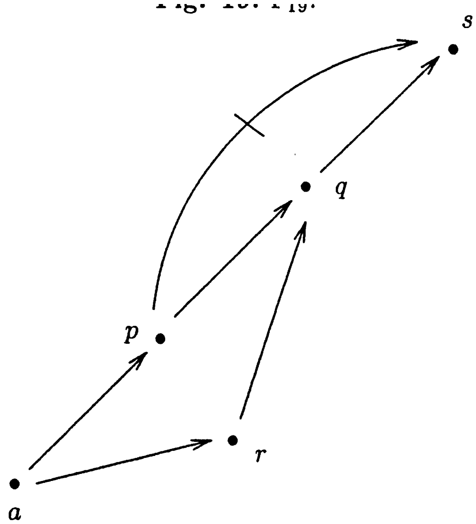

Fig. 20. Γ_{20}

to the detailed structure of the arguments through which statements are supported. It seems to have little to do with any ordinary idea of specificity, and is best motivated by thinking about argument paths in a more procedural way.

Suppose a path of the form $\pi(x_1, \ldots, x_n) \rightarrow y$ is constructible in the context $\langle \Gamma, \Phi \rangle$. In this case, because we have already accepted the argument $\pi(x_1, \ldots, x_n)$ telling us that x_1 is an x_n, the link $x_n \rightarrow y$ gives us a reason to accept the further conclusion that x_1 is a y. In the process of coming to accept the argument $\pi(x_1, \ldots, x_n)$, however, we would first have to accept each of its initial segments, telling us that x_1 is an x_2, an x_3, and so on. Now suppose $x_i \not\rightarrow y \in \Gamma$ for some $1 \leq i < n$. Then we cannot reach the point where $\pi(x_1, \ldots, x_n)$ gives us a reason for accepting the conclusion that x_1 is a y without first running into a reason for accepting instead the conclusion that x_1 is not a y. This illustrates the main idea behind on-path preemption: an argument path, which gives us a reason for accepting some conclusion, is supposed to be preempted whenever the path cannot even be constructed without first supplying a reason for a conflicting conclusion. If we can think of the reasons supplied by an argument path as determined not only by the nodes lying on that path, but also by all of its intermediaries, then this statement describes the on-path treatment of preemption exactly.

4.3 A fixed point approach to skeptical inheritance

The treatment of skeptical extensions presented earlier, in Section 2.2.2, relies crucially on the notion of degree. Unlike the credulous and flexible extensions, which are characterized through fixed point equations, the skeptical extensions are defined only through the iterative process of considering, and then accepting or rejecting, argument paths in the order of their degree. Because of this, it may seem that skeptical inheritance is fundamentally procedural in a way that credulous and flexible inheritance are not.

It was pointed out in Section 2.2.2 that the inheritance definition presented there does not quite coincide with the original definition of skeptical inheritance from [Horty *et al.*, 1990]. In fact, this original definition also relies on degree, also defining extensions through an iterative process; but as it turns out, the use of degree in that paper was not essential. The kind of skeptical extensions defined there could just as easily have been given through a fixed point equation. This section presents a fixed point approach to the skeptical extensions defined in [Horty *et al.*, 1990], and explains the difference between these and the extensions from Section 2.2.2 of the present chapter.

A key idea behind skeptical inheritance is the notion of a permitted path. In Section 2.2.2, the permitted paths are characterized using the notion of inheritability: a path is classified there as permitted if it is inheritable, but not opposed by any other inheritable path. In [Horty *et al.*, 1990], however, the permitted paths are defined in a way that bypasses the concept of inheritability, relying instead on a characterization of what it means for a path to be—let us say—*protected* in a context.

Definition 4.3.1 (Protection). A positive path $\pi(x, \sigma, u) \rightarrow y$ is *protected* in the context $\langle \Gamma, \Phi \rangle$ iff (i) $x \not\rightarrow y \notin \Gamma$ and (ii) for all nodes v such that $v \not\rightarrow y \in \Gamma$ and there is some path $\pi(x, \tau, v) \in \Phi$, there is a node z such that $z \rightarrow y \in \Gamma$ and either (a) $z = x$ or (b) there is some path $\pi(x, \tau_1, z, \tau_2, v) \in \Phi$. A negative path $\pi(x, \sigma, u) \not\rightarrow y$ is *protected* in the context $\langle \Gamma, \Phi \rangle$ iff (i) $x \rightarrow y \notin \Gamma$ and (ii) for all nodes v such that $v \rightarrow y \in \Gamma$ and there is some path $\pi(x, \tau, v) \in \Phi$, there is a node z such that $z \not\rightarrow y \in \Gamma$ and either (a) $z = x$ or (b) there is some path $\pi(x, \tau_1, z, \tau_2, v) \in \Phi$.

Using this idea of protection, we can set out an alternative definition of the permission relation, represented now by the symbol $\mid\!\sim'$.

Definition 4.3.2 (Permission: alternative definition).

Case I: σ is a direct link. Then $\langle \Gamma, \Phi \rangle \mid\!\sim' \sigma$ iff $\sigma \in \Gamma$.

Case II: σ is a compound path. Then $\langle \Gamma, \Phi \rangle \mid\!\sim' \sigma$ iff

1. σ is constructible in $\langle \Gamma, \Phi \rangle$,

2. σ is protected in $\langle \Gamma, \Phi \rangle$.

We can then define the skeptical extensions, alternatively, as the fixed points associated with this new relation of permission.

Definition 4.3.3. The path set Φ is a *skeptical extension (alternative definition)* of the net Γ iff

$$\Phi = \{\sigma : \langle \Gamma, \Phi \rangle \succ' \sigma \}.$$

And we can show that the fixed points associated with this relation, unlike the credulous or flexible extensions, are unique.

Theorem 4.3.4. *If Γ is an acyclic net, there is exactly one set Φ such that $\Phi = \{\sigma : \langle \Gamma, \Phi \rangle \succ' \sigma\}$.*

The skeptical extensions defined in this way coincide with the extensions defined in [Horty *et al.*, 1990]; and it is shown there that the analogs to Theorems 2.2.2, 2.2.3, and 2.2.4 hold for these extensions.

Now, how does the approach to skeptical inheritance adopted here, and in [Horty *et al.*, 1990], differ from that of Section 2.2.2? The root of the difference lies, of course, with the two notions of permission set out in Definitions 2.2.5 and 4.3.2. Although very similar, these two notions do not coincide. According to the earlier Definition 2.2.5, the permitted paths are a subset of the inheritable paths; and so a path that is preempted, and therefore not inheritable, cannot be permitted either. According to the present Definition 4.3.2, on the other hand, even a preempted path can be classified as permitted, if every path that preempts it is itself preempted. To see the effect of this distinction, consider the net Γ_{21} (Figure 21), for example; and imagine that the agent has reasoned his way to the context $\langle \Gamma_{21}, \Phi \rangle$, with $\Phi = \Gamma_{21} \cup \{x \rightarrow z \rightarrow v \rightarrow u\}$. According to Definition 2.2.5, the path $x \rightarrow z \rightarrow v \rightarrow u \rightarrow y$ cannot be permitted in this context, since it is preempted. According to Definition 4.3.2, however, the path will be permitted, since the path $x \rightarrow z \rightarrow v \not\rightarrow y$ which preempts it is itself preempted by the path $x \rightarrow z \rightarrow y$.

Because of the differences in their treatment of permitted paths, the two approaches to skeptical inheritance will at times lead to different extensions. In the case of Γ_{21}, again, the skeptical extension given by the current Definition 4.3.3 contains the path $x \rightarrow z \rightarrow v \rightarrow u \rightarrow y$, while that given by the earlier Definition 2.2.6 does not (the two extensions are otherwise identical). It may seem that this difference between these two extensions is not terribly significant, since after all, they differ only in the argument paths they contain, not in the statements they support: even according to Definition 2.2.6, the skeptical extension of Γ_{21} will support the

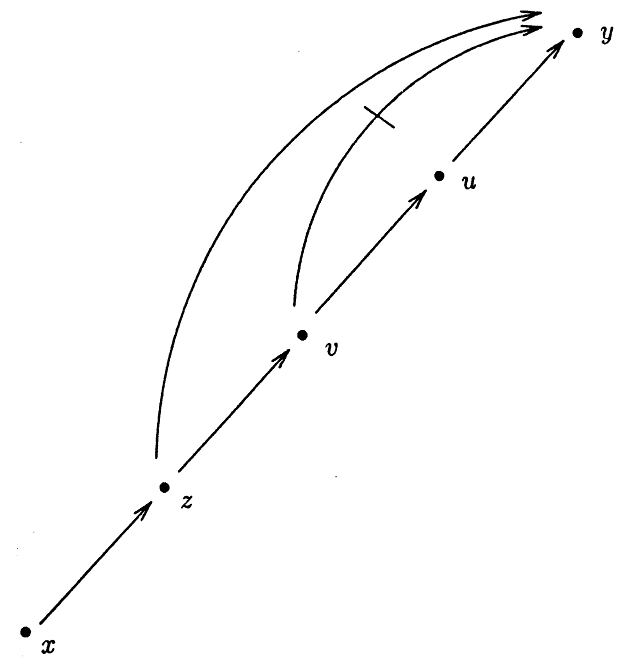

Fig. 21. Γ_{21}

statement $x \rightarrow y$, through the path $x \rightarrow z \rightarrow y$. As we have seen in Section 4.2, however, it is possible for the specificity relation that figures into preemption to be defined in a way that is sensitive, not only to the statements supported, but also to the paths through which they are supported. In particular, the off-path approach to preemption, which is incorporated both into our official definition of inheritability and into the current notion of protection, is one of those that determine specificity through paths rather than statements alone; and so we should expect to find cases in which differences in the paths through which a particular statement is supported give rise also to differences in the set of supported statements.

An example is provided by the net Γ_{22} (Figure 22). According to the treatment of Section 2.2.2, the skeptical extension of this net supports neither $a \rightarrow p$ not $a \not\rightarrow p$. According to the present treatment, however, the skeptical extension of this net contains the path $a \rightarrow r \not\rightarrow p$, and so supports $a \not\rightarrow p$. The reason for this variance is that the present treatment permits the path $a \rightarrow r \rightarrow s \rightarrow q$, so that it is possible to regard r as providing more specific information than q about a. On the previous treatment, however, this path is not permitted, even though the extension

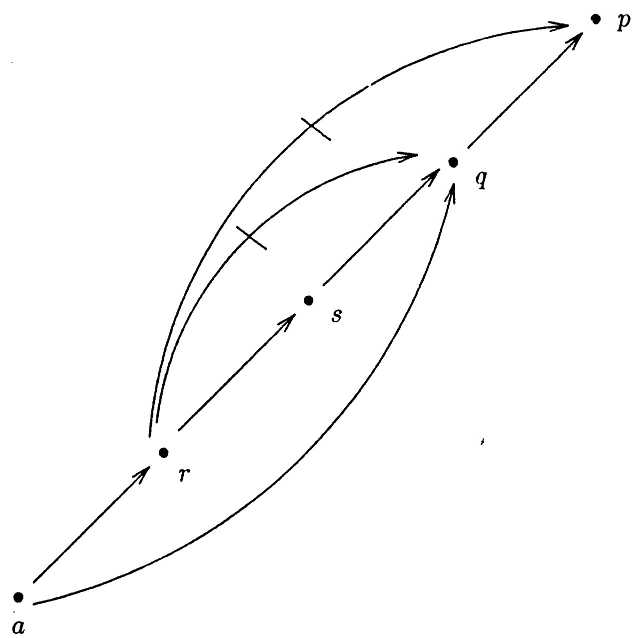

Fig. 22. Γ_{22}

nevertheless supports the statement $a \rightarrow q$ in some other way; and so a skeptical attitude is forced between $a \rightarrow r \not\rightarrow p$ and $a \rightarrow q \rightarrow p$.

A comparison between these two approaches to skeptical inheritance is contained in [Touretzky *et al.*, 1991].

4.4 Translational theories and meaning holism

We have concentrated in this chapter on direct theories of nonmonotonic inheritance, which attempt to specify the appropriate consequences of a network directly in terms of the network language itself. As noted earlier, however, there is also considerable interest in translational or indirect theories, those attempting to specify the consequences of a network by interpreting it in some more standard nonmonotonic formalism. This section describes some criteria of adequacy for such an interpretation. For the sake of concreteness, and because we cannot consider the full range of possible interpretations of networks into nonmonotonic logics, we focus on the particular problem of modeling credulous inheritance in ordinary default logic [Reiter, 1980]. As a translational task, this may seem especially straightforward, since credulous inheritance and default logic are so similar.

4.4.1 A mapping into default logic

Suppose that $\mathcal{T}$ is some general translation mapping each net Γ into a default theory $\mathcal{T}(\Gamma)$. Like the credulous accounts of inheritance, default logic associates with a theory, in general, multiple extensions. In order for $\mathcal{T}$ to count as an adequate translation, we must require some kind of correspondence between the credulous extensions of a network Γ and the extensions of the default theory $\mathcal{T}(\Gamma)$ into which it is translated. It is a just bit tricky to describe the correspondence exactly, however, because both network and default extensions can contain items foreign to the other. The extension of a net can contain paths supporting derived generic statements, but there is no way to derive new default rules in default logic. Each extension of a default theory will contain all the logical truths, but there is nothing like these in network extensions.

In defining a reasonable notion of correspondence between network and default extensions, therefore, we can require that they coincide only where they overlap in expressive power. Let L represent the set of literals belonging to the background predicate calculus underlying the default logic, where these literals include those of the network language. Of course, any default extension E determines a subset $E \cap L$ of the literals; and if Φ is a net extension, let us take $L(\Phi)$ to be the subset of these literals supported by Φ. (Example: if $\Phi = \{a \rightarrow p, p \rightarrow q \rightarrow r, a \rightarrow p \not\rightarrow q\}$, then $L(\Phi) = \{Pa, \neg Qa\}$.) These literal fragments represent the total extent of expressive overlap between the two kinds of extensions; and so we can require as a condition of correspondence only that these fragments agree. More exactly, where E is a network extension and Φ is an extension for some default theory, we will say that E and Φ correspond just in case $E \cap L = L(\Phi)$.

We can now formulate our first condition of adequacy for translation, a *correspondence* condition, in the following way: where Γ is a net and $\mathcal{T}(\Gamma)$ is its interpretation into default logic, there should be some net extension Φ of Γ corresponding to each default extension E of $\mathcal{T}(\Gamma)$, and there should be some default extension E of $\mathcal{T}(\Gamma)$ corresponding to each net extension Φ of Γ. The first of these clauses can be thought of as saying that the default interpretation of a net is sound with respect to credulous inheritance; the second that it is complete. It is not enough, however, to require for adequacy only that a translation $\mathcal{T}$ should map each network into a sound and complete interpretation in order to count as an intuitively adequate translation of nets into default logic. It seems appropriate to require also, as a second condition, that the translation should be *modular*, in the sense that $\mathcal{T}(\Gamma \cup \Delta) = \mathcal{T}(\Gamma) \cup \mathcal{T}(\Delta)$ for any nets Γ and Δ.[12] Ultimately, this modularity requirement means that the translation of a net into default logic can proceed link by link.

We will return shortly to the reasons underlying the modularity requirement; but let us first consider a couple of candidate translations from networks into default logic, in order to illustrate the difficulty of satisfying both of these two conditions at once.

The first translation—say, $\mathcal{T}_1$—is the most immediately obvious candidate. Where Γ is some network, $\mathcal{T}_1(\Gamma)$ is defined as the default theory $\langle W, D \rangle$ with

$$W = \{Pa : a \rightarrow p \in \Gamma\} \cup \{\neg Pa : a \not\rightarrow p \in \Gamma\},$$

$$D = \{(Px : Qx \;/\; Qx) : p \rightarrow q \in \Gamma\} \cup \{(Px : \neg Qx \;/\; \neg Qx) : p \not\rightarrow q \in \Gamma\}.$$

This translation is clearly modular; and it yields a sound and complete default interpretation for many simple nets, such as Γ_1 (the Nixon Diamond). Here, for example, $\mathcal{T}_1(\Gamma_1)$ is the default theory $\langle W, D \rangle$, where $W = \{Qa, Ra\}$ and $D = \{(Qx : Px \;/\; Px), (Rx : \neg Px \;/\; \neg Px)\}$. This theory has two extensions: $E_1 = Th(\{Qa, Ra, Pa\})$ and $E_2 = Th(\{Qa, Ra, \neg Pa\})$. The net Γ_1 itself has two credulous extensions: $\Phi_1 = \Gamma_1 \cup \{a \rightarrow q \rightarrow p\}$ and $\Phi_2 = \Gamma_1 \cup \{a \rightarrow q \not\rightarrow p\}$. And these different extensions stand in the proper kind of correspondence: $E_1 \cap L = L(\Phi_1)$ and $E_2 \cap L = L(\Phi_2)$.

The translation $\mathcal{T}_1$ works well, then, in the case of the Nixon Diamond. It works also (of course) for any net containing no conflicting paths at all, and (interestingly) for nets containing only diamond-like conflicts, no matter how deeply nested. However, as is well known, this simple translation fails to give the appropriate meaning to nets in which preemption comes into play.[13] In the case of Γ_2 (the Tweety Triangle), for example, $\mathcal{T}_1$ yields the default theory $\langle W, D \rangle$ where $W = \{Ra\}$ and $D = \{(Rx : Qx \;/\; Qx), (Qx : Px \;/\; Px), (Rx : \neg Px \;/\; \neg Px)\}$. This theory allows $Th(\{Ra, Qa, Pa\})$ as one of its extensions; but since there is no network extension of Γ_2 to which this default extension corresponds, the translation of this network is not sound.

One response to this problem—originally suggested in [Reiter and Criscuolo, 1981], and developed in detail by [Etherington and Reiter, 1983]—is to modify $\mathcal{T}_1$ in such a way that the reasons that might override the application of a rule are built explicitly into the logical representation of that rule. In the environment of Γ_2, for example, the link $q \rightarrow p$ would be

[12] Translations satisfying this condition are sometimes described as *local* [Ginsberg, 1990].

[13] Because default logic is built on top of a classical logic, it also fails in the case of nets containing conflicting literals, links of the form $a \rightarrow p$ and $a \not\rightarrow p$. But that is a different sort of problem, and we ignore it here.

translated, not by the normal default rule $(Qx : Px \;/\; Px)$, but instead by the semi-normal rule $(Qx : [Px \wedge \neg Rx] \;/\; Px)$. This new translation—call it $\mathcal{T}_2$—does provide a sound and complete representation of Γ_2 in default logic; when the normal translation of $q \rightarrow p$ is replace by its semi-normal variant, the anomalous extension allowed by $\mathcal{T}_1$ is avoided. However, $\mathcal{T}_2$ does not meet the standards set out here either, since it fails to satisfy the condition of modularity.

4.4.2 Modularity

Why is the modularity of a translational interpretation so important? If we think of a translational interpretation as an attempt to specify the meaning of a network—mapping the network language into a logical formalism for which a precise notion of consequence is already defined—then the root of the problem with non-modular translations is that they provide only a *holistic* account of meaning. Holism is the view, in linguistics and the philosophy of language, that meaning can properly be ascribed only to entire theories, not individual sentences—or that the meaning of an individual sentence can vary depending on the theory in which it is embedded. An expression of the view remarkably apt to the present context can be found in the following passage:

> What has experiential import is the corporate body of statements, and this import is not the simple sum of the experiential imports of the individual statements. In ordinary language as opposed to formalized language, this phenomenon is made even more pervasive by what is sometimes called the "nonmonotonicity" of the logic of everyday discourse .... If I say, "Hawks fly," I do *not* intend my hearer to deduce that a hawk with a broken wing will fly. What we expect depends on the whole network of beliefs. If language describes experience, it does so as a network, not sentence by sentence [Putnam, 1988, p. 9].

Of course, we are concerned here to specify the consequences of a net rather than its "experiential import," but if we read the passage with this substitution in mind, it provides a good description of the idea underlying non-modular translation. Although $\mathcal{T}_2(\Gamma_2)$ does seem to represent the meaning of network Γ_2 as a whole, it does not do so link by link. The rule $(Qx : [Px \wedge \neg Rx] \;/\; Px)$ does not represent the meaning of the link $q \rightarrow p$, since this link will be represented by another rule when it occurs in Γ_1, and by other rules still when it occurs in other nets. Instead of representing the meaning of this link, the rule captures, at best, the meaning of the link as it occurs in this particular net.

In the present environment, where the project is to provide a semantics for a network formalism, there are at least two reasons to avoid a holistic theory. The first of these is a problem about update that was pointed out in [Touretzky, 1984]. Any knowledge representation system should be able to accommodate updates in some natural fashion, but update can be a very complicated process if the meaning of representational items is holistic. If the translation of any particular link can depend on the entire network in which it occurs, then when one updates a net by adding a new link, it is not enough simply to update the translation of the net by supplementing it with the translation of this new link. Because of the addition of the new link, the overall environment in which the original links are embedded has now changed, and so translations of these links may have to be adjusted.

As an example, consider the interpretation assigned by $\mathcal{T}_2$ to the net Γ_1; preemption is not involved, and so it simply agrees with that of $\mathcal{T}_1$. Suppose, now, that we update this net by adding the new link $q \rightarrow r$. How should we construct the default theory that results from this net; what is $\mathcal{T}_2(\Gamma_1 \cup \{q \rightarrow r\})$? Because $\mathcal{T}_2$ is not modular, it is not enough simply to supplement $\mathcal{T}_2(\Gamma_1)$ with the translation of this new link—$(Qx : Rx \ / \ Rx)$. Instead, we have to realize that in the new environment the original interpretation of $r \rightarrow p$ is no longer adequate; because of the addition of the new link, the interpretation of this original link must now be changed to $(Rx : [\neg Px \wedge \neg Qx] \ / \ \neg Px)$. Although the update operation defined on the net was quite simple—just a matter of adding a new link—the corresponding operation on the default theory is more complicated.

A second reason for avoiding non-modular, or holistic, interpretations of network formalisms can be seen by thinking about the original point of supplying a semantic analysis for these systems. Originally, we needed the analysis because, at least in certain cases, it was hard to see exactly what the networks meant, what conclusions they ought to support; and the semantics was supposed to tell us that. In order for a translational interpretation to do this job, however, the task of mapping the network formalism into the language for which the semantics is actually defined must be routine and mechanical; and in particular, the mapping cannot itself rely on insights about the meaning of the network, since that is what the semantics is supposed to provide.

A modular translation is the ideal. Here, the translation of a link in a network cannot possibly rely on insights about the meaning of that network, since the link has the same translation no matter where it occurs. Non-modular strategies, however, tend to require a fairly sophisticated understanding of the meaning of a net simply to achieve the proper translation; and in the case of theories like default logic, where the priorities among conflicting defaults that represent preemption relations in network must be coded explicitly, there is a very general reason for this. As we have seen in Section 4.2, the preemption relations in a network can themselves depend on the results of defeasible reasoning (which itself may be influenced by other preemption relations, themselves established through defeasible reasoning, and so on). In order for the preemption relations from a particular network to be encoded properly into a default theory, therefore, all of the defeasible reasoning necessary to establish these relations must be carried out *before* the translation takes place. This kind of reasoning, of course, can be arbitrarily difficult; and it can embody all the problematic features that suggested in the first place the need for a precise semantic theory to specify the meaning of network.

To illustrate this point about the dependence of non-modular translation on prior defeasible reasoning, consider the interpretation using semi-normal defaults even of a simple net like the earlier Γ_{17}—where, as we have seen, the priority relations among rules vary from extension to extension. In fact, all of the generic links in this net can be represented through normal defaults, except for the link $q \rightarrow r$. How should we translate this link? Should it be represented simply as $(Qx : Rx \ / \ Rx)$, which would place it on a par with the conflicting link $p \not\rightarrow r$, or should it be represented as $(Qx : [Rx \wedge \neg Px] \ / \ Rx)$, which would give the other link priority. The answer is that neither of these candidates is correct. According to any of the treatments of preemption that rely on defeasible specificity, the rule $p \not\rightarrow r$ is supposed to be preferred to $q \rightarrow r$ only in those extensions that contain the path $p \rightarrow t \rightarrow u$ instead of the conflicting $p \rightarrow s \not\rightarrow u$. In order to capture the intended meaning of the link $q \rightarrow r$ in this environment, therefore, we must represent it through a rule like $(Qx : [Rx \wedge (Ux \supset \neg Px)] \ / \ Rx)$, giving priority to the link $p \not\rightarrow r$ only for those objects that we have decided are u's.

With this rule in place, the translation of Γ_{17} into default logic does provide a sound and complete interpretation of this net. The translation can be said to give us the meaning of the net, but only in a very weak sense. Since the interpretation is non-modular, we could not just translate the individual links in a routine way, and then let the logic tell us what conclusions were contained in its different extensions. Instead, we had to have a pretty firm idea already of the conclusions contained in its different extensions to begin with, simply in order to translate simply net properly. We did not use the translation to discover the network's meaning; we discovered its meaning through some other kind of reasoning, and then simply designed the translation to yield the right results.

4.4.3 Other translations

Although we have focused here on a particular strategy for mapping inheritance networks into a particular logic, the kind of problems described arise more generally. For example, Etherington [Etherington, 1987; Etherington, 1988] suggests a way of modifying the mapping into default logic in order to meet Touretzky's criticisms; but the modified mapping seems also to run afoul of the modularity requirement. Likewise, it may seem possible to code inheritance networks into prioritized circumscriptive theories [Lifschitz, 1985], but it hard to see how this could be done in a modular way. In the case of circumscription, the problem does not lie with translation of the particular links, which is uniform, but with the prioritization of abnormalities in the circumscription schema. Again, these priorities may have to be rearranged as the knowledge base is updated; and again, the arrangement must be based on pretheoretic intuitions. We cannot use the meaning ascribed to a net by prioritized circumscription to get the priorities, since we need the priorities before the theory can be applied.

With only a few exceptions, such as [McCarty and Cohen, 1990], researchers working in this area seem now to have accepted the modularity requirement as a criterion of adequacy for a logical interpretation of inheritance networks. There have recently been a number of proposals for interpreting networks into more expressive nonmonotonic formalisms that address this requirement, at least; and several are claimed to satisfy it.

Perhaps the most interesting from a mathematical perspective is the interpretation in [Gelfond and Przymusinska, 1990] of inheritance networks into autoepistemic logic; this paper contains a number of illuminating theorems and a useful discussion of the principles underlying inheritance reasoning from a more general point of view. A different kind of mapping, into a simplified default logic, is set out in [Ginsberg, 1990]; this interpretation is especially provocative because, unlike the direct theories described here, and unlike most other logical interpretations, it allows contrapositive reasoning based on defeasible information.

Interpretations of inheritance networks within the framework of logic programming are set out in [Grégoire, 1989a; Grégoire, 1989b] and in [Lin, 1991]. Grégoire relies on theories of negation for stratified logic programs, while Lin appeals only to the simple negation as failure rule. Grégoire's work illustrates the points emphasized here about the need for a pretheoretic analysis of preemption to guide logical interpretation. He derives the priorities that condition his interpretation of a particular network from an application to that network of a direct theory of skeptical inheritance; and so it can be argued that it is this direct theory that constitutes the core of his analysis. However, Grégoire is explicit about what he is doing, and he supplies some interesting arguments to justify the strategy; the basic idea is that a translation of inheritance nets into the language of logic programming or default logic can be valuable even if it does depend on the results of some direct theory, because it can then be used to explain the meaning of the inheritance net to person or system that understands only logic programming or default logic.

An interpretation of inheritance based on circumscription is set out in [Haugh, 1988], which develops some earlier ideas from [McCarthy, 1986] Another interpretation based on circumscription is presented in [Krishnaprasad *et al.*, 1989a]. This work, along with [Krishnaprasad and Kifer, 1989] and [Krishnaprasad *et al.*, 1989b], is based on results obtained in [Krishnaprasad, 1989].

Another interesting interpretation is that of [Padgham, 1989b], which interprets inheritance networks within a lattice of concepts supplemented with additional operators allowing for the representation, for example, of typical instances of those concepts; portions of this work have appeared as [Padgham, 1988] and [Padgham, 1989a].

There have recently been developed several more generally expressive nonmonotonic formalisms that, like inheritance theories, embody implicit preferences between conflicting defaults, often based on specificity (or logical strength) of the antecedent. These include the argument systems of [Loui, 1987], [Nute, 1988], and [Pollock, 1987]; the preference based conditional logics of [Kraus *et al.*, 1990]; the conditional logics of [Boutilier, 1992] and [Delgrande, 1987; Delgrande, 1988]; and a logic proposed in [Geffner, 1989] that is motivated by probabilistic considerations. To date, none of these logics has been applied in any general analysis of inheritance network; but it would be very interesting to compare the independently motivated preference criteria developed in these theories with those at work in the direct approaches to inheritance. A slightly different way of relating inheritance networks to more expressive logics with implicit preference among defaults is explored in [Horty, 1992]. Here, instead of interpreting nets into a previously existing logic, it is shown that the ideas found in direct inheritance theories can themselves be generalized to apply to a full logical language; the perspective of the paper is that IS-A inheritance can then be thought of simply as the implicational fragment of this language.

Finally, a number of authors have recently investigated probabilistic and statistical interpretations of the information contained in defeasible inheritance networks. For example, [Bacchus, 1989] uses a simple probabilistic approach to define a system, unlike any of those considered here, in which chaining can proceed over only a single defeasible link. Pearl's ϵ-semantics is deployed in [Geffner and Pearl, 1987] to interpret positive and negative defeasible links as expressing probabilities arbitrarily close to one or zero; the interpretation supports chaining over multiple defeasible links. And [Neufeld, 1991] deploys a probabilistic interpretation to discuss some of the issues raised in [Touretzky *et al.*, 1987].

4.5 Implementational concerns

As explained in the introduction, this chapter has presented inheritance theory from a biased standpoint, emphasizing conceptual issues over the kind of concerns that might arise from an application of the work to knowledge representation. As a partial remedy to this bias, we conclude simply by providing some pointers to the literature on a couple of the concerns most relevant to applications—complexity and expressiveness.

4.5.1 Algorithms and complexity

The properties of algorithms for nonmonotonic inheritance reasoning have been a matter of concern since the subject was first introduced. Even the very early [Minsky, 1974] was motivated by the idea that the incorporation of default information into a knowledge base could simplify the process of information retrieval. And Fahlman's dissertation [Fahlman, 1979], which was inspired by Minsky's work, contained an informal argument that, in ordinary cases, the parallel inference algorithms described there could be expected to perform in time linear in the depth of the knowledge base.

On the basis of his formal reconstruction of Fahlman's ideas into a credulous theory of inheritance, Touretzky was able to show that that Fahlman's argument was correct in special cases, but problematic for networks allowing multiple extensions. As a remedy, he proposed a technique by which, given an extension, a network could be conditioned in such a way that algorithms much like those described by Fahlman would lead to the appropriate results in linear time; this work is described in [Touretzky, 1986, Chapter 4], which also introduces a notation to specify the marker propagation algorithms envisaged by Fahlman for carrying out inheritance reasoning on a parallel machine.

In the case of skeptical inheritance, tractability for the definition presented here in Section 4.3, and originally in [Horty *et al.*, 1990], was established in that paper.[14] The paper describes a parallel marker propagation algorithm that answers queries of the form "Is x a y?" in a time roughly equal to $O(D(x,y) + C(x,y))$, where $D(x,y)$ is the depth of the network once it is "trimmed" so that links irrelevant to the query at hand are ignored, and and $C(x,y)$ is the number of conflicts contained in the trimmed network. In realistic knowledge bases, where conflicts are expected to be infrequent, the time for this query algorithm approaches the purely parallel $O(D(x,y))$, which approximates Fahlman's original intuition.

A different approach to skeptical inheritance, discussed here in Sections 2.2.1 and 2.2.2, identifies the skeptically supported conclusions of a network with those conclusions supported by each of its credulous extensions. If we put aside the matter of preemption, then the results concerning normal default theories from [Kautz and Selman, 1991] can be seen as providing a polynomial algorithm for computing skeptical inheritance in this sense. Subsequently, [Selman, 1990] showed that adding almost any form of preemption leads to intractability for this form of skep-

[14] The algorithm there relies on certain idiosyncratic features of the inheritance definition contained in that paper that are missing from the definition presented here in Section 2.2.2. However, Selman (personal communication) has established the tractability of the definition from Section 2.2.2.

tical reasoning. One exception, however, is the reasoner of [Stein, 1989; Stein, forthcoming], which describes a very interesting labeling algorithm for computing the intersection of credulous extensions that incorporates some notion of preemption and does run in polynomial time.

It is shown in [Selman and Levesque, 1989] that, for a variety of path-based inheritance definitions, including at least those of [Touretzky, 1986] and [Horty *et al.*, 1990], the question of tractability hinges on the kind of chaining involved in path construction. Definitions based on the standard, upward chaining notion of path construction presented here in Section 2.1 lead to tractable query algorithms; those definitions based on the alternative, double chaining notion presented in Section 4.1 lead to query algorithms that are NP-complete. These surprising results, which are described more fully in [Selman, 1990], show that the issues discussed in Section 4.1 from a conceptual point of view have important computational consequences as well.

All of the formal work on complexity mentioned so far concerns purely defeasible inheritance networks. Inference in strict IS-A networks is trivial; a purely parallel algorithm that draws conclusions in accord with the definition of [Thomason *et al.*, 1986] is provided there. However, as usual, the matter becomes much more complicated with strict and defeasible information are mixed together, and little is currently known about the complexity of reasoning in this case. Some preliminary results in [Touretzky and Thomason, 1990] indicate that a low order polynomial algorithm is still available; but the particular algorithm presented there was not verified for correctness, and has revealed some problems subsequent to publication. It is established in [Goldszmidt and Pearl, 1991], however, that at least the consistency of a mixed net can be established rather easily—in only $O(E^2)$ time, where E is the number of links in the net.

4.5.2 Expressive enhancements

In fact, there have been only a few applications within actual knowledge representation systems of the kind work described in this chapter on the semantics of nonmonotonic inheritance reasoning. There are, of course, a variety of knowledge representations systems that incorporate monotonic inheritance (including KL-ONE and its descendants), and any number of systems relying on nonmonotonic inheritance of an unprincipled kind; but the applications of principled nonmonotonic inheritance have been sparse.[15]

[15] One example is the work of [Rector, 1986], who has built a sizable knowledge base of medical information organized as an inheritance hierarchy; the query algorithms he defines rely on preemption to handle exceptions to inherited information, but there are severe limitations on the structure of the knowledge base. Another example is the recent use of nonmonotonic inheritance in representing grammatical information; a description along with references appears in [Thomason, 1991].

A major reason for this is that most of the defeasible formalisms for which semantic theories have actually been developed are extremely limited in their expressive resources. Fahlman originally envisaged in [Fahlman, 1979] a richly expressive network formalism. In order to focus on the semantic problems presented by nonmonotonic inheritance in a constrained environment, however, Touretzky limited his attention in [Touretzky, 1986] only to the defeasible IS-A fragment of Fahlman's rich formalism; and most of the theoretical research on nonmonotonic inheritance that has appeared since then has dealt only with this limited language or very modest extensions.

For the purposes of realistic knowledge representation, of course, the restriction to IS-A links is insufficient: there is just not much that can be said in such a limited language. In retrospect, it seems to have been a good idea to focus research initially on this simple fragment, since so many interesting problems can be seen so clearly there. But the past few years have done much to articulate our semantic understanding of these problems; and at present, the most important research objective in the theoretical study of nonmonotonic inheritance should be to broaden the scope of coverage to include languages with expressive resources more suitable to realistic knowledge representation. Some work along these lines has already started. For example, as a prelude to incorporating defeasible information into these structures, [Thomason and Touretzky, 1992] investigates the theory of strict networks containing roles and relations. And it is shown in [Horty and Thomason, 1990] that the kind of theories described in the present chapter can be extended to provide a semantics for defeasible networks containing complex nodes defined through boolean operations.

Acknowledgments

My greatest debt in writing this chapter is to Rich Thomason and Dave Touretzky: most of what I know about nonmonotonic inheritance I learned through discussion or collaborative work with them, and their influence throughout the chapter is pervasive. In addition, I have received useful comments on an earlier draft from a number of people, but especially from Matt Ginsberg, Moisés Goldszmidt, Bart Selman, and Lynn Stein.

This work is supported by the National Science Foundation under Grant No. IRI-9003165.

A Proofs of selected theorems

Theorem 2.2.8. *Φ is a credulous extension of an acyclic net Γ iff Φ is the limit of a credulous reasoning sequence based on Γ.*

Proof. The proof has two parts.

Part 1. Let $\Phi_1, \Phi_2, \ldots$ be a credulous reasoning sequence based on Γ. We show that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$ is a credulous extension of Γ.

Part 1A: if $\sigma \in \Phi$ then $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$. Suppose $\sigma \in \Phi$. If σ is a link, the definition of the sequence tells us that $\sigma \in \Gamma$; and so $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$. Next, let σ be a compound path, with $\deg_\Gamma(\sigma) = n + 1$. Then the definition of the reasoning sequence tells us that $\sigma \in \Phi_{n+1}$, and so $\langle \Gamma, \Phi_n \rangle \mid\!\sim \sigma$—that is, σ is constructible but neither conflicted nor preempted in the context $\langle \Gamma, \Phi_n \rangle$. Since σ is constructible in $\langle \Gamma, \Phi_n \rangle$, it is plainly constructible also in $\langle \Gamma, \Phi \rangle$. To show that $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$, we show that σ is neither conflicted nor preempted in this context.

Suppose Φ contains a path τ conflicting with σ. Then $\deg_\Gamma(\tau) = \deg_\Gamma(\sigma)$, and so $\tau \in \Phi_{n+1}$. Moreover, the definition of the reasoning sequence tells us that both σ and τ are first introduced at Φ_{n+1}; that is, both σ and τ belong to $\Phi_{n+1} - \Phi_n$. But this is impossible, since $\Phi_{n+1} - \Phi_n$ must be a conflict free set.

It is easy to see that σ cannot be preempted in $\langle \Gamma, \Phi \rangle$ either. Let σ be a positive path of the form $\pi(x, \sigma_1, u) \to y$ (the argument for negative paths is similar), and suppose it is preempted in $\langle \Gamma, \Phi \rangle$. Then there is a node v such that $v \not\to y \in \Gamma$ and either $v = x$ or $\pi(x, \tau_1, v, \tau_2, u) \in \Phi$. Since $\deg_\Gamma(\sigma) = n+1$, we know that $\deg_\Gamma(\pi(x, \tau_1, v, \tau_2, u)) = \deg_\Gamma(\pi(x, \sigma_1, u)) \leq n$. The definition of the reasoning sequence then tells us that either $v = x$ or $\pi(x, \tau_1, v, \tau_2, u) \in \Phi_n$, and so σ must be preempted already in $\langle \Gamma, \Phi_n \rangle$, contrary to assumption.

Part 1B: if $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$ then $\sigma \in \Phi$. Suppose $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$. If σ is a link, then $\sigma \in \Gamma$; but $\Gamma = \Phi_1 \subseteq \Phi$. Next, let σ be a compound path. Then σ is constructible but neither conflicted nor preempted in the context $\langle \Gamma, \Phi \rangle$. Suppose σ is a positive path of the form $\pi(x, \sigma_1, u) \to y$ (the argument for negative paths is similar), and that $\deg_\Gamma(\sigma) = n + 1$.

Since σ is constructible in $\langle \Gamma, \Phi \rangle$, we know that $\pi(x, \sigma_1, u) \in \Phi$ and $u \to y \in \Gamma$. However, $\deg_\Gamma(\pi(x, \sigma_1, u)) \leq n$, and so the definition of the reasoning sequence tells us that $\pi(x, \sigma_1, u) \in \Phi_n$. So σ is constructible also in $\langle \Gamma, \Phi_n \rangle$. Moreover, σ cannot be conflicted or preempted in $\langle \Gamma, \Phi_n \rangle$, since it is not conflicted or preempted in $\langle \Gamma, \Phi \rangle$, and $\Phi_n \subseteq \Phi$. Therefore $\langle \Gamma, \Phi_n \rangle \mid\!\sim \sigma$, and so $\sigma \in I_{n+1}(\Gamma, \Phi_n)$. From this, we can conclude that $\sigma \in \Phi_{n+1} \subseteq \Phi$. For suppose $\sigma \notin \Phi_{n+1}$. Then since $\Phi_{n+1} - \Phi_n$ is a maximal conflict free subset of $I_{n+1}(\Gamma, \Phi_n)$, there must be some $\tau \in \Phi_{n+1} - \Phi_n$ that conflicts with σ. But then $\tau \in \Phi$, and so σ is conflicted in $\langle \Gamma, \Phi \rangle$, contrary to assumption.

Part 2. Let Φ be a credulous extension of Γ. We show that there exists a credulous reasoning sequence $\Phi_1, \Phi_2, \ldots$ based on Γ such that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$. The sequence is defined by taking $\Phi_1 = \Gamma$ and

$$\Phi_{n+1} = \Phi_n \cup \{\sigma \in \Phi : \deg_\Gamma(\sigma) = n + 1\}.$$

Since it is clear that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$, we need only show that $\Phi_1, \Phi_2, \ldots$ is a credulous reasoning sequence; and for this it is enough to establish that $\Xi = \{\sigma \in \Phi : \deg_\Gamma(\sigma) = n+1\}$ is a maximal conflict free subset of $I_{n+1}(\Gamma, \Phi_n)$.

We show first that Ξ is a subset of $I_{n+1}(\Gamma, \Phi_n)$. Take $\sigma \in \Xi$. Since we know from the definition of Ξ that $\deg_\Gamma(\sigma) = n+1$, it is necessary only to show that $\langle \Gamma, \Phi_n \rangle \mathrel{|\!\sim} \sigma$. Since $\Xi \subseteq \Phi$, we have $\sigma \in \Phi$; and so $\langle \Gamma, \Phi \rangle \mathrel{|\!\sim} \sigma$, since Φ is an extension. Let σ have the form $\pi(x, \sigma_1, u) \to y$ (negative paths are similar). Since σ is constructible in $\langle \Gamma, \Phi \rangle$, we know that $\pi(x, \sigma_1, u) \in \Phi$ and $u \to y \in \Gamma$. But $\deg_\Gamma(\pi(x, \sigma_1, u)) \leq n$, and so the definition of the sequence tells us that $\pi(x, \sigma_1, u) \in \Phi_n$. Therefore σ is constructible also in $\langle \Gamma, \Phi_n \rangle$, and since $\Phi_n \subseteq \Phi$, it is neither conflicted nor preempted. So $\langle \Gamma, \Phi_n \rangle \mathrel{|\!\sim} \sigma$.

Next we show that Ξ is conflict free. Suppose it contains two conflicting paths. Since $\Xi \subseteq \Phi$, both of these paths must belong to Φ; but then Theorem 2.2.3 tells us that the paths must both be direct links, in which case they would both have a degree of 1, and could not belong to Ξ.

Finally, we show that Ξ is maximal among the conflict free subsets of $I_{n+1}(\Gamma, \Phi_n)$. Suppose $\sigma \in I_{n+1}(\Gamma, \Phi_n)$, but that $\sigma \notin \Xi$. Since $\sigma \in I_{n+1}(\Gamma, \Phi_n)$, we know that σ is inheritable—constructible, but neither conflicted nor preempted—in the context $\langle \Gamma, \Phi_n \rangle$. Since $\sigma \notin \Xi$, it follows that $\sigma \notin \Phi$; and so since Φ is an extension, we know that σ is not inheritable in the context $\langle \Gamma, \Phi \rangle$. Because σ is constructible in $\langle \Gamma, \Phi_n \rangle$, it must be constructible also in $\langle \Gamma, \Phi \rangle$. Moreover, an argument based on degree similar to that in Part 1A of this proof shows that σ cannot be preempted in $\langle \Gamma, \Phi \rangle$. Therefore, σ must be conflicted in this context; there must be some path $\tau \in \Phi$ that conflicts with σ. But in this case, we have $\deg_\Gamma(\tau) = \deg_\Gamma(\sigma)$, and so $\tau \in \Xi$. Hence Ξ is maximal. ■

Theorem 2.2.9. *Φ is a flexible extension of an acyclic net Γ iff Φ is the limit of a flexible reasoning sequence based on Γ.*

Proof. The proof has two parts.

Part 1. Let $\Phi_1, \Phi_2, \ldots$ be a flexible reasoning sequence based on Γ. We show that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$ is a flexible extension of Γ.

Part 1A: if $\sigma \in \Phi$ then $\langle \Gamma, \Phi \rangle \mathrel{|\!\succ} \sigma$. Suppose $\sigma \in \Phi$. If σ is a link, the definition of the sequence tells us that $\sigma \in \Gamma$; and so $\langle \Gamma, \Phi \rangle \mathrel{|\!\succ} \sigma$. Next, let σ be a compound path, with $\deg_\Gamma(\sigma) = n+1$. Then the definition of the sequence tells us that $\sigma \in \Phi_{n+1}$, and so $\langle \Gamma, \Phi_n \rangle \mathrel{|\!\succ} \sigma$—that is, we have

(1) $\langle \Gamma, \Phi_n \rangle \mathrel{|\!\sim} \sigma$,

(2) there is no path τ such that $\langle \Gamma, \Phi_n \rangle \mathrel{|\!\sim} \tau$ and τ conflicts with σ.

From (1) we can conclude that

(1′) $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$,

by an argument analogous to that in Part 1A of the proof of Theorem 2.2.8. Now suppose that there is some path τ conflicting with σ such that $\langle \Gamma, \Phi \rangle \mid\!\sim \tau$. By an argument similar to that in Part 1B of the proof of Theorem 2.2.8, we can then conclude that $\langle \Gamma, \Phi_n \rangle \mid\!\sim \tau$; but this contradicts (2). Therefore we have

(2′) there is no path τ such that $\langle \Gamma, \Phi \rangle \mid\!\sim \tau$ and τ conflicts with σ;

and so $\langle \Gamma, \Phi \rangle \mid\!\!\!- \sigma$.

Part 1B: if $\langle \Gamma, \Phi \rangle \mid\!\!\!- \sigma$ *then* $\sigma \in \Phi$. Suppose $\langle \Gamma, \Phi \rangle \mid\!\!\!- \sigma$. If σ is a link, then $\sigma \in \Gamma$; but $\Gamma = \Phi_1 \subseteq \Phi$. Next, let σ be a compound path, with $\deg(\sigma) = n + 1$. Then we have

(1) $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$,

(2) there is no path τ such that $\langle \Gamma, \Phi \rangle \mid\!\sim \tau$ and τ conflicts with σ.

From (1) we can conclude by an argument similar to that in Part 1B of the proof of Theorem 2.2.8 that $\langle \Gamma, \Phi_n \rangle \mid\!\sim \sigma$, so that $\sigma \in I_{n+1}(\Gamma, \Phi_n)$. Now there are two cases to consider. First, suppose that $I_{n+1}(\Gamma, \Phi_n)$ contains no argument conflicting with σ. Then it follows at once from the definition of the reasoning sequence that $\sigma \in \Phi_{n+1} \subseteq \Phi$. Next, suppose there is an argument $\tau \in I_{n+1}(\Gamma, \Phi_n)$ that conflicts with σ. Then we know $\langle \Gamma, \Phi_n \rangle \mid\!\sim \tau$—that is, τ is constructible, but nether conflicted nor preempted in $\langle \Gamma, \Phi_n \rangle$—but (2) tells us that we cannot have $\langle \Gamma, \Phi \rangle \mid\!\sim \tau$. Since τ is constructible in $\langle \Gamma, \Phi_n \rangle$, it must be constructible also in $\langle \Gamma, \Phi \rangle$; and an argument similar to that in Part 1A of the proof of Theorem 2.2.8 allows us to conclude that τ cannot be preempted in $\langle \Gamma, \Phi \rangle$. Therefore, τ must be conflicted in $\langle \Gamma, \Phi \rangle$; there must be some argument σ' in Φ that conflicts with τ. Since σ' conflicts with τ, we have $\deg_\Gamma(\sigma') = \deg_\Gamma(\tau) = \deg_\Gamma(\sigma) = n + 1$. So $\sigma' \in \Phi_{n+1}$. Moreover, since σ' and σ both conflict with τ, they have to support the same statement. Therefore, since $\Phi_{n+1} - \Phi_n$ is a statement uniform subset of $I_{n+1}(\Gamma, \Phi_n)$, we must have $\sigma \in \Phi_{n+1} \subseteq \Phi$ as well.

Part 2. Let Φ be a flexible extension of Γ. We show that there exists a flexible reasoning sequence $\Phi_1, \Phi_2, \ldots$ based on Γ such that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$. As before,the sequence is defined by taking $\Phi_1 = \Gamma$ and

$$\Phi_{n+1} = \Phi_n \cup \{\sigma \in \Phi : \deg_\Gamma(\sigma) = n + 1\}.$$

Since it is clear that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$, we need only show that $\Phi_1, \Phi_2, \ldots$ is a flexible reasoning sequence; and for this it is enough to establish that

$\Xi = \{\sigma \in \Phi : \deg_\Gamma(\sigma) = n+1\}$ is a conflict free and statement uniform subset of $I_{n+1}(\Gamma, \Phi_n)$.

The arguments that Ξ is a subset of $I_{n+1}(\Gamma, \Phi_n)$ and that it is conflict free are identical with the analogous arguments in Part 2 of the proof of Theorem 2.2.8; so we need only show that Ξ is statement uniform. Suppose that $\sigma_1, \sigma_2 \in I_{n+1}(\Gamma, \Phi_n)$, where these two paths support the same statement. In that case, of course, we have $\deg_\Gamma(\sigma_1) = \deg_\Gamma(\sigma_2) = n+1$, and both $\langle \Gamma, \Phi_n \rangle \mid\!\sim \sigma_1$ and $\langle \Gamma, \Phi_n \rangle \mid\!\sim \sigma_2$—both paths are constructible in this context, but neither conflicted nor preempted. The paths σ_1 and σ_2 must then be constructible also in $\langle \Gamma, \Phi \rangle$, and an argument like that in Part 1A of the proof of Theorem 2.2.8 shows that they cannot be preempted there either. Moreover, since these paths support the same statement, each will be conflicted in $\langle \Gamma, \Phi \rangle$ only if the other is; and so we have $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma_1$ iff $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma_2$. Again, since these paths support the same statement, any path τ such that $\langle \Gamma, \Phi \rangle \mid\!\sim \tau$ conflicting with either will conflict also with the other. Therefore, we have $\langle \Gamma, \Phi \rangle \mid\!\succ \sigma_1$ iff $\langle \Gamma, \Phi \rangle \mid\!\succ \sigma_2$. Since Φ is a flexible extension, it follows from this that $\sigma_1 \in \Phi$ iff $\sigma_2 \in \Phi$. From this we can conclude also that $\sigma_1 \in \Xi$ iff $\sigma_2 \in \Xi$; and so Ξ is statement uniform. ∎

Theorem 3.3.3. *Φ is a credulous extension of Γ iff Φ is the limit of a credulous mixed reasoning sequence for Γ.*

Proof. The proof has two parts.

Part 1. Let $\Phi_1^0, \Phi_1^1, \Phi_2^0, \Phi_2^1, \ldots$ be a credulous mixed reasoning sequence based on Γ. We show that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n^1$ is a credulous extension of Γ.

Part 1A: if $\sigma \in \Phi$ then $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$. Suppose $\sigma \in \Phi$. We consider four cases, depending on the structure of σ. f σ is a defeasible link, the definition of the reasoning sequence tells us that $\sigma \in \Gamma$, and so we know from Case C-I of the inheritability definition (Definition 3.2.3) that $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$. If σ is a strict path, the definition of the reasoning sequence tells us that σ is constructible from the links of Γ and so we know from Case B of the inheritability definition that $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$.

Next, suppose σ is a compound defeasible path that does not possess a strict end segment. Then $\mathrm{mdeg}_\Gamma(\sigma) = \langle n+1, 0 \rangle$; so the definition of the reasoning sequence tells us that $\sigma \in \Phi_{n+1}^0$; and so $\langle \Gamma, \Phi_n^1 \rangle \mid\!\sim \sigma$—that is, σ is constructible but neither conflicted nor preempted in $\langle \Gamma, \Phi_n^1 \rangle$. Since σ is constructible in $\langle \Gamma, \Phi_n^1 \rangle$, it is plainly constructible also in $\langle \Gamma, \Phi \rangle$. We show that σ is neither conflicted nor preempted in this context, from which it follows by Case C-II of the inheritability definition that $\langle \Gamma, \Phi \rangle \mid\!\sim \sigma$.

Suppose Φ contains a path τ conflicting in Γ with σ. Note first that the defeasible degree of τ must agree with that of σ, since any two compound, defeasible paths that conflict in a mixed net must have the same defeasible degree in that net. Now suppose τ has no strict end segment. Then $\mathrm{mdeg}_\Gamma(\tau) = \langle n+1, 0 \rangle$, and so $\tau \in \Phi_{n+1}^0$. Moreover, the definition of the reasoning sequence tells us that both σ and τ are first introduced at Φ^0_{n+1}; that is, both σ and τ belong to $\Phi^0_{n+1} - \Phi^1_n$. But this is impossible, since $\Phi^0_{n+1} - \Phi^1_n$ must be a conflict free set. Now suppose τ does possess a strict end segment. Then $\text{mdeg}_\Gamma(\tau) = \langle n+1, 1\rangle$; so the definition of the reasoning sequence tells us that $\tau \in \Phi^1_{n+1}$; and so $\langle \Gamma, \Phi^0_{n+1}\rangle \mathrel{|\!\!\sim} \tau$. From this we can conclude by Case A of the inheritability definition that $\text{Def}(\tau) \in \Phi^0_{n+1}$; and of course, $\text{mdeg}_\Gamma(\text{Def}(\tau)) = \langle n+1, 0\rangle$, so that $\text{Def}(\tau)$ is first introduced at Φ^0_{n+1}. It follows from the definition of conflict (Definition 3.2.1), however, that $\text{Def}(\tau)$ itself conflicts with σ, since τ does. Therefore, $\text{Def}(\tau)$ cannot be introduced at Φ^0_{n+1}, since $\Phi^0_{n+1} - \Phi^1_n$ must be conflict free.

It is easy to see that σ cannot be preempted in $\langle \Gamma, \Phi\rangle$ either. Let σ be a positive path of the form $\pi(x, \sigma_1, u) \to y$ (the argument for negative paths is similar), and suppose it is preempted in $\langle \Gamma, \Phi\rangle$. Then there exist nodes v, m such that $v = x$ or $\pi(x, \tau_1, v, \tau_2, u) \in \Phi$ and either (a) $v \to m \in \Gamma$ and $m \in \overline{\kappa}(y)$ or (b) $v \not\to m \in \Gamma$ and $m \in \kappa(y)$. It is clear, however, that $\text{mdeg}_\Gamma(\pi(x, \tau_1, v, \tau_2, u)) = \text{mdeg}_\Gamma(\pi(x, \sigma_1, u)) \leq \langle n, 1\rangle$. Therefore, either $v = x$ or $\pi(x, \tau_1, v, \tau_2, u) \in \Phi^1_n$; and so σ is preempted already in $\langle \Gamma, \Phi^1_n\rangle$, contrary to assumption.

Finally, suppose σ is a compound defeasible path that does possess a strict end segment. Then $\text{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle$; so the definition of the reasoning sequence tells us that $\sigma \in \Phi^1_n$; and so $\langle \Gamma, \Phi^0_n\rangle \mathrel{|\!\!\sim} \sigma$—that is $\text{Def}(\sigma) \in \Phi^0_n$ and $\text{Str}(\sigma) \in \Phi^0_n$. However, since $\Phi^0_n \subseteq \Phi$, Case A of Definition 3.2.3 then tells us that $\langle \Gamma, \Phi\rangle \mathrel{|\!\!\sim} \sigma$ as well.

Part 1B: if $\langle \Gamma, \Phi\rangle \mathrel{|\!\!\sim} \sigma$ *then* $\sigma \in \Phi$. Suppose $\langle \Gamma, \Phi\rangle \mathrel{|\!\!\sim} \sigma$. If σ is a defeasible path that does not possess a strict end segment, then it can be shown that $\sigma \in \Phi$ through an argument similar to that contained in Part 1B of the proof of Theorem 2.2.8. If σ is a strict path, Case B of the inheritability definition tells us that σ is constructible from the links in Γ; and then the definition of the reasoning sequence allows us to conclude that $\sigma \in \Phi^0_1 \subseteq \Phi$.

Finally, suppose σ is a defeasible path that does possess a strict end segment, where $\text{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle$. Then Case A of the inheritability definition tells us that $\text{Def}(\sigma) \in \Phi$ and $\text{Str}(\sigma) \in \Phi$. We know $\text{mdeg}_\Gamma(\text{Def}(\sigma)) = \langle n, 0\rangle$; so the definition of the reasoning sequence tells us that $\text{Def}(\sigma) \in \Phi^0_n$; and of course, $\text{Str}(\sigma) \in \Phi^0_n$ also, since $\Phi^0_1 \subseteq \Phi^0_n$. Therefore, $\langle \Gamma, \Phi^0_n\rangle \mathrel{|\!\!\sim} \sigma$; and so the definition of the reasoning sequence tells us that $\sigma \in \Phi^1_n \subseteq \Phi$.

Part 2. Let Φ be a credulous extension of Γ. We show that there exists a credulous mixed reasoning sequence $\Phi^0_1, \Phi^1_1, \Phi^0_2, \Phi^1_2, \ldots$ based on Γ such that $\Phi = \bigcup_{n=1}^{\infty} \Phi^1_n$. The sequence is defined by setting $\Phi_1 = \Gamma \cup \{\sigma :$ $\text{Str}(\sigma) = \sigma$ and σ constructed from $\Gamma\ \}$, and then taking

$$\begin{array}{lclcl} \Phi^1_n & = & \Phi^0_n & \cup & \{\sigma \in \Phi : \text{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle\}, \\ \Phi^0_{n+1} & = & \Phi^1_n & \cup & \{\sigma \in \Phi : \text{mdeg}_\Gamma(\sigma) = \langle n+1, 0\rangle\}. \end{array}$$

Since it is clear that $\Phi = \bigcup_{n=1}^{\infty} \Phi_n$, we need only show that $\Phi^0_1, \Phi^1_1, \Phi^0_2, \Phi^1_2, \ldots$ is a credulous reasoning sequence. For this, it is enough to establish (1) that $\Omega = \{\sigma \in \Phi : \text{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle\}$ is identical to $I_n^1(\Gamma, \Phi_n^0)$, and (2) that $\Xi = \{\sigma \in \Phi : \deg_\Gamma(\sigma) = \langle n+1, 0\rangle\}$ is a maximal conflict free subset of $I_{n+1}(\Gamma, \Phi_n)$. We prove here only the first of these claims; the argument for the second is similar to that presented in Part 2 of the proof of Theorem 2.2.8.

We show first that Ω is a subset of $I_n^1(\Gamma, \Phi_n^0)$. Suppose $\sigma \in \Omega$. Since we know that $\text{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle$, it is enough to show that $\langle \Gamma, \Phi_n^0\rangle \mid\!\sim \sigma$. Since $\Omega \subseteq \Phi$, we have $\sigma \in \Phi$; and so $\langle \Gamma, \Phi\rangle \mid\!\sim \sigma$, since Φ is an extension. From this, it follows by Case A of the inheritability definition that $\text{Def}(\sigma) \in \Phi$ and $\text{Str}(\sigma) \in \Phi$. Clearly, $\text{mdeg}_\Gamma(\text{Def}(\sigma)) = \langle n, 0\rangle$, so that $\text{Def}(\sigma) \in \Phi_n^0$; and of course $\text{Str}(\sigma) \in \Phi_n^0$, since $\Phi_n^0 \subseteq \Phi_1^0$. Therefore, $\langle \Gamma, \Phi_n^0\rangle \mid\!\sim \sigma$ by Case A of inheritability.

Next we show that $I_n^1(\Gamma, \Phi_n^0)$ is a subset of Ω. Suppose $\sigma \in I_n^1(\Gamma, \Phi_n^0)$. Since we know that $\text{mdeg}_\Gamma(\sigma) = \langle n, 1\rangle$, it is enough to show that $\sigma \in \Phi$. Because $\sigma \in I_n^1(\Gamma, \Phi_n^0)$, we know also that $\langle \Gamma, \Phi_n^0\rangle \mid\!\sim \sigma$; so $\text{Def}(\sigma) \in \Phi_n^0$ and $\text{Str}(\sigma) \in \Phi_n^0$, by Case A; so $\text{Def}(\sigma) \in \Phi$ and $\text{Str}(\sigma) \in \Phi$, since $\Phi_n^0 \subseteq \Phi$; and so $\langle \Gamma, \Phi\rangle \mid\!\sim \sigma$, by Case A. From this is follows that $\sigma \in \Phi$, since Φ is an extension. ∎

Theorem 4.3.4. *If* Γ *is an acyclic net, there is exactly one set* Φ *such that* $\Phi = \{\sigma : \langle \Gamma, \Phi\rangle \mid\!\!\sim' \sigma\}$.

Proof. Suppose both $\Phi_1 = \{\sigma : \langle \Gamma, \Phi_1\rangle \mid\!\sim' \sigma\}$ and $\Phi_2 = \{\sigma : \langle \Gamma, \Phi_2\rangle \mid\!\sim' \sigma\}$. We show by induction on $\deg_\Gamma(\sigma)$ that $\sigma \in \Phi_1$ iff $\sigma \in \Phi_2$. If $\deg_\Gamma(\sigma) = 1$, then σ is a link; so $\sigma \in \Gamma$, and so $\langle \Gamma, \Phi\rangle \mid\!\sim' \sigma$ for any Φ. Next suppose as inductive hypothesis that $\tau \in \Phi_1$ iff $\tau \in \Phi_2$ whenever $\deg_\Gamma(\tau) < \deg_\Gamma(\sigma)$

Take $\sigma \in \Phi_1$, where σ has the form $\pi(x, \sigma_1, u) \rightarrow y$ (the negative case is similar). Since $\sigma \in \Phi_1$, we know $\langle \Gamma, \Phi_1\rangle \mid\!\sim' \sigma$, so that σ is both constructible and protected in the context $\langle \Gamma, \Phi_1\rangle$. Since σ is constructible, we have $\pi(x, \sigma_1, u) \in \Phi_1$ and $u \rightarrow y \in \Gamma$. But $\deg_\Gamma(\pi(x, \sigma_1, u)) < \deg_\Gamma(\sigma)$. So $\pi(x, \sigma_1, u) \in \Phi_2$ by induction, and so σ is constructible also in $\langle \Gamma, \Phi_2\rangle$.

Now suppose σ is not protected in $\langle \Gamma, \Phi_2\rangle$. Then either (i) $x \not\rightarrow y \in \Gamma$ or (ii) there is a node v for which $v \not\rightarrow y \in \Gamma$ and there is some path $\pi(x, \tau, v) \in \Phi_2$, but no node z for which $z \rightarrow y \in \Gamma$ and either (a) $z = x$ or (b) there is some path $\pi(x, \tau_1, z, \tau_2, v) \in \Phi_2$. If (i), then σ cannot be protected in $\langle \Gamma, \Phi_1\rangle$, contrary to assumption. So suppose (ii). Then since we have both $\deg_\Gamma(\pi(x, \tau, v)) < \deg_\Gamma(\sigma)$ and $\deg_\Gamma(\pi(x, \tau_1, z, \tau_2, v)) < \deg_\Gamma(\sigma)$, the inductive hypothesis tells us that (ii) holds also for Φ_1; so again, σ cannot be protected in $\langle \Gamma, \Phi_1\rangle$, contrary to assumption. Therefore, σ is protected in Φ_2, and so we have $\langle \Gamma, \Phi_2\rangle \mid\!\sim' \sigma$. Therefore $\sigma \in \Phi_2$. ∎

References

[Anderson and Belnap, 1975] Alan Anderson and Nuel Belnap. *Entailment: The Logic of Relevance and Necessity.* Princeton University Press, 1975.

[Bacchus, 1989] Fahiem Bacchus. A modest, but semantically well founded, inheritance reasoner. In *Proceedings of IJCAI-89*, pages 1104–1109. Morgan Kaufmann, 1989.

[Barwise, 1977] Jon Barwise, editor. *Handbook of Mathematical Logic.* North Holland, 1977.

[Belnap, 1977a] Nuel Belnap. How a computer should think. In Gilbert Ryle, editor, *Contemporary Aspects of Philosophy*, pages 30–56. Oriel Press, 1977.

[Belnap, 1977b] Nuel Belnap. A useful four-valued logic. In J. Michael Dunn and G. Epstein, editors, *Modern Uses of Multiple-valued Logic*, pages 8–37. D. Reidel, 1977.

[Bobrow and Winograd, 1977] Daniel Bobrow and Terry Winograd. An overview of KRL, a knowledge representation language. *Cognitive Science*, 1:3–46, 1977.

[Boutilier, 1989a] Craig Boutilier. On the semantics of stable inheritance reasoning. Technical Report KRR-TR-89-11, Department of Computer Science, University of Toronto, 1989. An abbreviated version appears as [Boutilier, 1989b].

[Boutilier, 1989b] Craig Boutilier. On the semantics of stable inheritance reasoning. In *Proceedings of IJCAI-89.* Morgan Kaufmann, 1989.

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

[Brachman and Levesque, 1985] Ronald Brachman and Hector Levesque, editors. *Readings in Knowledge Representation.* Morgan Kaufmann, 1985.

[Brachman, 1985] Ronald Brachman. "I lied about the trees," or definitions and defaults in knowledge representation. *The AI Magazine*, 6(Fall):80–93, 1985.

[Delgrande, 1987] James Delgrande. A first-order conditional logic for prototypical properties. *Artificial Intelligence*, 33:105–130, 1987.

[Delgrande, 1988] James Delgrande. An approach to default reasoning based on a first-order conditional logic: revised report. *Artificial Intelligence*, 36:63–90, 1988.

[Doyle, 1985] Jon Doyle. Reasoned assumptions and Pareto optimality. In *Proceedings of IJCAI-85*, pages 87–90, 1985.

[Doyle, 1988] Jon Doyle. On universal theories of defaults. Technical Report CMU-CS-88-111, Computer Science Department, Carnegie Mellon University, 1988.

[Etherington and Reiter, 1983] David Etherington and Raymond Reiter. On inheritance hierarchies with exceptions. In *Proceedings of AAAI-83*, pages 104–108, 1983.

[Etherington, 1987] David Etherington. Formalizing nonmonotonic reasoning systems. *Artificial Intelligence*, 31:41–85, 1987.

[Etherington, 1988] David Etherington. *Reasoning with Incomplete Information.* Morgan Kaufmann, 1988.

[Fahlman *et al.*, 1981] Scott Fahlman, David Touretzky, and W. van Roggen. Cancellation in a parallel semantic network. In *Proceedings of IJCAI-81*, pages 257–263, 1981.

[Fahlman, 1979] Scott Fahlman. *NETL: A System for Representing and Using Real-world Knowledge.* The MIT Press, 1979.

[Gabbay, 1985] Dov Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In Krzysztof Apt, editor, *Logics and Models of Concurrent Systems*, pages 439–457. Springer Verlag, 1985.

[Geffner and Pearl, 1987] Hector Geffner and Judea Pearl. Sound defeasible inference. Technical Report CSD-870058, Computer Science Department, University of California at Los Angeles, 1987.

[Geffner and Verma, 1989] Hector Geffner and Tom Verma. Inheritance = chaining + defeat. In Zbigniew Ras, editor, *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*, pages 411–418. North Holland, 1989.

[Geffner, 1989] Hector Geffner. *Default Reasoning: Causal and Conditional Theories.* PhD thesis, Computer Science Department, University of California at Los Angeles, 1989.

[Gelfond and Przymusinska, 1990] Michael Gelfond and Helena Przymusinska. Formalization of inheritance reasoning in autoepistemic logic. *Fundamenta Informaticae*, 13:403–443, 1990.

[Gentzen, 1934] Gerhard Gentzen. Untersuchungen über das logische schliessen. *Mathematische Zeitschrift*, 39:176–210, 1934. English translation (as "Investigations into logical deduction") in [Szabo, 1969], pages 68–131.

[Ginsberg, 1990] Matthew Ginsberg. A local formalization of inheritance: preliminary report. Manuscript, Computer Science Department, Stanford University, 1990.

[Goldszmidt and Pearl, 1991] Moisés Goldszmidt and Judea Pearl. On the consistency of defeasible databases. *Artificial Intelligence*, pages 121–149, 1991.

[Grégoire, 1989a] Eric Grégoire. *Logiques non monotones, Programmes logiques stratifié et Théories sceptiques de l'héritage*. PhD thesis, Université de Louvain, Belgium, 1989.

[Grégoire, 1989b] Eric Grégoire. Skeptical theories of inheritance and nonmonotonic logics. In *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*, pages 430–438. North Holland, 1989.

[Haugh, 1988] Brian Haugh. Tractable theories of multiple defeasible inheritance in ordinary nonmonotonic logics. In *Proceedings of AAAI-88*, pages 421–426. Morgan Kaufmann, 1988.

[Hayes, 1977] Patrick Hayes. In defense of logic. In *Proceedings of IJCAI-77*, pages 87–90, 1977.

[Hayes, 1979] Patrick Hayes. The logic of frames. In D. Metzing, editor, *Frame Conceptions and Text Understanding*, pages 46–61. Walter de Gruyter and Co., 1979. Reprinted in [Brachman and Levesque, 1985], pages 287–297.

[Horty and Thomason, 1988] John Horty and Richmond Thomason. Mixing strict and defeasible inheritance. In *Proceedings of AAAI-88*, pages 427–432. Morgan Kaufmann, 1988.

[Horty and Thomason, 1990] John Horty and Richmond Thomason. Boolean extensions of inheritance networks. In *Proceedings of AAAI-90*, pages 633–639. AAAI Press and MIT Press, 1990.

[Horty *et al.*, 1987] John Horty, Richmond Thomason, and David Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. In *Proceedings of AAAI-87*, pages 358–363, 1987.

[Horty *et al.*, 1990] John Horty, Richmond Thomason, and David Touretzky. A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence*, 42:311–348, 1990. A preliminary version appears as [Horty *et al.*, 1987].

[Horty, 1991] John Horty. A credulous theory of mixed inheritance. In Maurizio Lenzerini, Daniele Nardi, and Maria Simi, editors, *Inheritance*

Hierarchies in Knowledge Representation and Programming Languages. John Wiley and Sons, 1991.

[Horty, 1992] John Horty. Defeasible arguments: a logical extension of path-based inheritance reasoning. Manuscript, Institute for Advanced Computer Studies, University of Maryland, 1992.

[Horty, forthcoming] John Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, forthcoming.

[Kautz and Selman, 1991] Henry Kautz and Bart Selman. Hard problems for simple default logic. *Artificial Intelligence*, 49:243–279, 1991.

[Kraus *et al.*, 1990] Sarit Kraus, Daniel Lehman, and Menachem Magidor. Nonmonotonic reasoning, preferential models, and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Krishnaprasad and Kifer, 1989] T. Krishnaprasad and Michael Kifer. An evidence based theory of inheritance. In *Proceedings of IJCAI-89*, pages 1093–1098. Morgan Kaufmann, 1989.

[Krishnaprasad *et al.*, 1989a] T. Krishnaprasad, Michael Kifer, and David Warren. On the circumscriptive semantics of inheritance networks. In *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems*, pages 448–456. North Holland, 1989.

[Krishnaprasad *et al.*, 1989b] T. Krishnaprasad, Michael Kifer, and David Warren. On the declarative semantics of inheritance networks. In *Proceedings of IJCAI-89*, pages 1098–1103. Morgan Kaufmann, 1989.

[Krishnaprasad, 1989] T. Krishnaprasad. *The Semantics of Inheritance Networks*. PhD thesis, Computer Science Department, State University of New York at Stony Brook, 1989.

[Lifschitz, 1985] Vladimir Lifschitz. Computing circumscription. In *Proceedings of IJCAI-85*, pages 121–127, 1985.

[Lin, 1991] Fangzhen Lin. *A Study in Nonmonotonic Reasoning*. PhD thesis, Computer Science Department, Stanford University, 1991.

[Loui, 1987] Ronald Loui. Defeat among arguments: A system of defeasible inference. *Computational Intelligence*, 3:100–106, 1987.

[Makinson and Schlechta, 1991] David Makinson and Karl Schlechta. Floating conclusions and zombie paths: two deep difficulties in the "directly skeptical" approach to defeasible inheritance nets. *Artificial Intelligence*, 48:199–209, 1991.

[Makinson, 1989] David Makinson. General theory of cumulative inference. In Michael Reinfrank, Johan de Kleer, Matthew Ginsberg, and

Eric Sandewall, editors, *Proceedings of the Second International Workshop on Nonmonotonic Reasoning*, volume 346 of *Springer Verlag Lecture Notes in Artificial Intelligence*, pages 1–18. Springer Verlag, 1989.

[McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[McCarty and Cohen, 1990] L. Thorne McCarty and W. Cohen. The case for explicit exceptions. Manuscript, Computer Science Department, Rutgers University, 1990.

[McDermott, 1978] Drew McDermott. Tarskian semantics, or no notation without denotation! *Cognitive Science*, 2:277–282, 1978.

[Minsky, 1974] Marvin Minsky. A framework for representing knowledge. Technical Report 306, Artificial Intelligence Laboratory, MIT, 1974. Reprinted in [Brachman and Levesque, 1985], pages 246–262.

[Neufeld, 1991] Eric Neufeld. Notes on 'A clash of intuitions'. *Artificial Intelligence*, 48:225–240, 1991.

[Nute, 1988] Donald Nute. Defeasible reasoning: A philosophical analysis in prolog. In J. Fetzer, editor, *Aspects of Artificial Intelligence*, pages 251–288. Kluwer Academic Publishers, 1988.

[Padgham, 1988] Lin Padgham. A model and representation for type information and its use in reasoning with defaults. In *Proceedings of AAAI-88*, pages 409–414. Morgan Kaufmann, 1988.

[Padgham, 1989a] Lin Padgham. Negative reasoning using inheritance. In *Proceedings of IJCAI-89*. Morgan Kaufmann, 1989.

[Padgham, 1989b] Lin Padgham. *Non-Monotonic Inheritance for an Object-Oriented Knowledge-Base*. PhD thesis, Department of Computer and Information Science, Linköping University, 1989.

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

[Prawitz, 1971] Dag Prawitz. Ideas and results in proof theory. In J. Fensted, editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 235–307. North Holland, 1971.

[Prawitz, 1972] Dag Prawitz. The philosophical position of proof theory. In *Contemporary Logic in Scandinavia*. The Johns Hopkins Press, 1972.

[Putnam, 1988] Hilary Putnam. *Representation and Reality*. The MIT Press, 1988.

[Rector, 1986] A. Rector. Defaults, exceptions and ambiguity in a medical knowledge representation system. *Medical Informatics*, 11:295–306, 1986.

[Reiter and Criscuolo, 1981] Raymond Reiter and Giovanni Criscuolo. On interacting defaults. In *Proceedings of IJCAI-81*, pages 270–276, 1981.

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

[Roberts and Goldstein, 1977] R. Bruce Roberts and Ira Goldstein. The FRL manual. Technical Report AI Memo No. 409, Artificial Intelligence Laboratory, MIT, 1977.

[Sandewall, 1986] Eric Sandewall. Non-monotonic inference rules for multiple inheritance with exceptions. *Proceedings of the IEEE*, 74:1345–1353, 1986.

[Selman and Levesque, 1989] Bart Selman and Hector Levesque. The tractability of path-based inheritance. In *Proceedings of IJCAI-89*. Morgan Kaufmann, 1989.

[Selman, 1990] Bart Selman. *Tractable Defeasible Reasoning*. PhD thesis, Computer Science Department, University of Toronto, 1990.

[Stein, 1989] Lynn Stein. Skeptical inheritance: Computing the intersection of credulous extensions. In *Proceedings of IJCAI-89*. Morgan Kaufmann, 1989.

[Stein, 1990] Lynn Stein. *Resolving Ambiguity in Nonmonotonic Reasoning*. PhD thesis, Computer Science Department, Brown University, 1990.

[Stein, forthcoming] Lynn Stein. Resolving ambiguity in nonmonotonic inheritance hierarchies. *Artificial Intelligence*, forthcoming.

[Szabo, 1969] M.E. Szabo, editor. *The Collected Papers of Gerhard Gentzen*. North Holland, 1969.

[Thomason and Touretzky, 1992] Richmond Thomason and David Touretzky. Inheritance theory and networks with roles. In John Sowa, editor, *Principles of Semantic Networks*, pages 231–266. Morgan Kaufmann, 1992.

[Thomason *et al.*, 1986] Richmond Thomason, John Horty, and David Touretzky. A calculus for inheritance in monotonic semantic nets. Technical Report CMU-CS-86-138, Computer Science Department, Carnegie

Mellon University, 1986. An abbreviated version appears as [Thomason *et al.*, 1987].

[Thomason *et al.*, 1987] Richmond Thomason, John Horty, and David Touretzky. A calculus for inheritance in monotonic semantic nets. In Zbigniew Ras and Maria Zemankova, editors, *Proceedings of the Second International Symposium on Methodologies for Intelligent Systems*, pages 280–287. North-Holland, 1987.

[Thomason, 1991] Richmond Thomason. Inheritance and natural language. Manuscript, Intelligent Systems Program, University of Pittsburgh, 1991.

[Thomason, 1992] Richmond Thomason. NETL and subsequent path-based inheritance theories. *Computers and Mathematics with Applications*, 23:179–204, 1992.

[Touretzky and Thomason, 1990] David Touretzky and Richmond Thomason. An inference algorithm for networks that mix strict and defeasible inheritance. In Z. Ras, editor, *Proceedings of the Fifth International Symposium on Methodologies for Intelligent Systems.* North Holland, 1990.

[Touretzky *et al.*, 1987] David Touretzky, John Horty, and Richmond Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Proceedings of IJCAI-87*, pages 476–482. Morgan Kaufmann, 1987.

[Touretzky *et al.*, 1991] David Touretzky, Richmond Thomason, and John Horty. A skeptic's menagerie: conflictors, preemptors, reinstaters, and zombies in nonmonotonic inheritance. In *Proceedings of IJCAI-91*, pages 478–483. Morgan Kaufmann, 1991.

[Touretzky, 1984] David Touretzky. Implicit ordering of defaults in inheritance systems. In *Proceedings of AAAI-84*, pages 322–325. Morgan Kaufmann, 1984.

[Touretzky, 1986] David Touretzky. *The Mathematics of Inheritance Systems.* Morgan Kaufmann, 1986.

[Woods, 1975] William Woods. What's in a link: foundations for semantic networks. In Daniel Bobrow and A. Collins, editors, *Representation and Understanding: Studies in Cognitive Science*, pages 35–82. Academic Press, 1975. Reprinted in [Brachman and Levesque, 1985], pages 218–241.