Inheritance Reasoning: Psychological Plausibility, Proof Theory and Semantics

Carl Vogel

Doctor of Philosophy
University of Edinburgh
1995

Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated.

Carl M. Vogel

Edinburgh, 25 April 1995

Abstract

Default inheritance reasoning is a propositional approach to nonmonotonic reasoning designed to model reasoning with natural language generics. Inheritance reasoners model sets of natural language generics as directed acyclic graphs, and inference corresponds to the specification of paths through those networks. A proliferation of inheritance proof theories exist in the literature along with extensive debate about the most reasonable way to construct inferences, based on intuitions about interpretations of particular inheritance networks. There has not been an accepted semantics for inheritance which unifies the set of possible proof theories, which would help identify truly ill-motivated proof theories. This thesis attempts to clarify the inheritance literature in the three ways indicated in the title: psychological plausibility, proof theory and semantics. The thesis intends to displace debate about the best inferences to draw about a network from logicians' introspections to empirical investigation of how people respond to sets of defaults. The third chapter investigates a range of assumptions made within the literature on default inheritance reasoning and additionally attempts to arbitrate among conflicting inheritance reasoners based on the best fit with empirical data. This empirical research is among the first studies which have attempted to inform inheritance proof theory with clues about the way people reason. The second and fourth chapters of the thesis contribute to inheritance proof theory by identifying complexity results that have been misleading, and by defining some new inheritance systems based in part by the observations made in analyzing the data accumulated in experiments reported in Chapter Three. The main contribution of the proof theoretic chapters is the presentation of a set of definitions for a range of inheritance reasoners within a well defined family of path-based systems. It is shown how the reasoners relate to each other in terms of both the conclusions reached as well as in indicating the changes to the definitions required to obtain the appropriate mutations. Implementations are supplied for a number of these systems. The fifth chapter uses this parameterization of inheritance proof theory in providing a unified semantics using channel theory, a new mathematical framework designed to model natural regularities and information flow. The essential modeling tool supplied by channel theory is the information channel, an object in the universe that enables information flow and which stands for regularities. The semantics for inheritance uses information channels as the interpretation of inheritance links. Inferences, corresponding to paths through the inheritance networks, are interpreted by composite channels. Thus, various reasoning strategies have a natural interpretation in corresponding restrictions on the putative serial composition of channels. This is the first unified semantics for a family of path-based inheritance reasoners. The thesis concludes by pointing out the various ways in which this work can be extended.

Acknowledgements

I would like to express deep gratitude to Jeff Horty, Richmond Thomason and David Touretzky for developing such a thick area of nonmonotonic reasoning to explore in a PhD thesis. Thanks to Robin Cooper, for giving me freedom to wallow in this stuff as I pleased, and for giving guidance and advice at quite crucial moments, support at even more crucial ones. I'd also like to thank Elisabet Engdahl for additional sound advice through the course of my studies in Edinburgh. Marc Moens has also been quite helpful, more than he knows, by giving me things to read and letting me just get on with it. Lawrence Cavedon, Tsutomu Fujinami, Claire Hewson, Ulrike Hahn and Catherine Collin were all very generous in making time to read and comment on various stages of this material. I also owe thanks to both Jeff Pelletier and Jon Oberlander for agreeing to formally examine the thesis.

I am still completely tickled that it is the Centre for Cognitive Science here in Edinburgh where I've been able to do my PhD, and the Marshall Aid Commemoration Commision has my eternal gratitude for making it possible (as well as Robin's EKN boxes for letting it continue to be). Special thanks to Betty Hughes for making administration of all that quite painless, and for keeping the main office so warm & friendly. A lot of people have made Edinburgh a fun experience from a social point of view: Pat Healey, Claire Hewson, Roisín Higgins (I don't think it would have been possible to finish this thesis within the desired timetable if she hadn't been planning to submit a PhD thesis on almost exactly the same day and available for commiseration through late-license pints), Lex Holt, Tomaž Erjavec, Jelena Meznaric, Sheila Rock, Andreas Schöter, and Paul Schweizer are the prime culprits there. Particular thanks are due to Pat and Tomaž and especially Claire for putting up with me in collaborations and sundry other ways that are best left unspoken here. I'd quite like to thank whoever invented e-mail for making it possible to keep in such frequent communication with folks across the water – Fred Popowich, Nick Cercone, Marcus Smith, Tony Lopez, Glenn MacDonald, Joerg Ueberla, Chris Wiseman, my parents and Dianna Laurent, my sister and brainstorming partner. It has been really good of my family to still acknowledge my existence when I pass through Louisiana after such lengthy absences, as do so many close friends (Tim & Michelle Murphy, Beth Campbell, Michele Barrere, ...) who have long given up on me answering letters by post in a timely fashion.

Finally, I'd like to dedicate this thesis to John James Boepple, my best friend from childhood, who was murdered in New Orleans on July 21, 1994. This thesis prevented me from giving him a eulogy at his funeral, so I reckon it belongs to him.

Contents

List of Figures

Chapter 1

Introduction

1.1 Defeasible Inference

This thesis is about the combination of facts that birds fly, that penguins are birds, that penguins do not fly, and that people are not in the least bothered by this state of affairs. However, since the workings of most computer systems are based on classical two-valued logics, this presents a dilemma if they are to be used to reason like or about people. The classical first order representation of these statements is given in (1.1).

$$\begin{aligned}
&\forall x,\ \mathit{bird}(x) \supset \ \mathit{flier}(x) \\
&\forall x,\ \mathit{penguin}(x) \supset \ \mathit{bird}(x) \\
&\forall x,\ \mathit{penguin}(x) \supset \neg\, \mathit{flier}(x)
\end{aligned} \tag{1.1}$$

On their own, those sentences are a bit alarming given the implied contradiction, but in a first order system they lead only to the conclusion that penguins do not exist. However, given an additional fact as in (1.2) there is a contradiction that cannot be explained away. In a classical system a contradiction is fatal, since although the system becomes instantly decidable (and in fact, with constant time complexity in proving any statement true or false), the system actually becomes useless since any sentence at all is thus trivially provable.

$$
\begin{array}{ll}
 & \forall x,\ \mathit{bird}(x) \supset\ \mathit{flier}(x) \\
 & \forall x,\ \mathit{penguin}(x) \supset\ \mathit{bird}(x) \\
 & \forall x,\ \mathit{penguin}(x) \supset \neg\ \mathit{flier}(x) \\
 & \exists Opus,\ \mathit{penguin}(Opus) \\
\cline{2-2}
(1.2) & \therefore - \\
 & \therefore \exists Elvis \\
 & \therefore \smile \\
 & \therefore \ldots
\end{array}
$$

It is difficult to patch first order logic to cope with the acquisition of knowledge that is not compatible with what is already known.

$$
(1.3) \quad
\begin{array}{l}
\forall x,\ \mathit{bird}(x) \supset\ \mathit{flier}(x) \\
\forall x,\ \mathit{penguin}(x) \supset\ \mathit{bird}(x) \\
\exists x,\ \mathit{penguin}(x)
\end{array}
$$

If (1.3) represents what is known at some particular time, and this is subsequently augmented by the fact that $\forall x,\ \mathit{penguin}(x) \supset \neg\ \mathit{flier}(x)$, then at the very least some meta-system is required to examine each known fact and modify the set until a stable state is reached again as in the consistent first order representation in (1.4). For any nontrivial domain, this will clearly be quite a complex process since there are so many individual exceptions.

$$
(1.4) \quad
\begin{array}{l}
\forall x,\ \mathit{bird}(x) \wedge \neg\ \mathit{penguin}(x) \supset\ \mathit{flier}(x) \\
\forall x,\ \mathit{penguin}(x) \supset\ \mathit{bird}(x) \\
\forall x,\ \mathit{penguin}(x) \supset \neg\ \mathit{flier}(x) \\
\exists x,\ \mathit{penguin}(x)
\end{array}
$$

Of course, it is possible to abstract over exceptions by defining classes of predicates for each of them (as in "birds which are not known to not fly, fly"). However, this just yields a more aesthetic representation of each fact since antecedents do not have to list each individual sort of exception; it remains as complex to perform an update to the set of known facts (for a survey of meta-logical approaches, see

Genesereth & Nilsson, 1988).

$$
(1.5)\qquad
\begin{aligned}
&\forall x,\ \mathit{bird}(x) \wedge \neg\ \mathit{abnormal1}(x) \supset\ \mathit{flier}(x)\\
&\forall x,\ \mathit{penguin}(x) \supset\ \mathit{abnormal1}(x)\\
&\forall x,\ \mathit{ostrich}(x) \supset\ \mathit{abnormal1}(x)\\
&\forall x,\ \mathit{penguin}(x) \supset\ \mathit{bird}(x)\\
&\forall x,\ \mathit{penguin}(x) \supset \neg\ \mathit{flier}(x)\\
&\exists x,\ \mathit{penguin}(x)
\end{aligned}
$$

An alternative approach is to adopt a modal logic which adds an operator C which performs a consistency check over the entire corpus of known facts. Thus, (1.6) also means, "birds which are not known to not fly, fly", but achieves that in a more concise fashion than listing exceptions directly or through exception predicates; however, it is at the expense of computing logical closure of the entire knowledge base.

$$
(1.6)\qquad \forall x,\ \mathit{bird}(x) \wedge\ C(\ \mathit{flier}(x)) \supset\ \mathit{flier}(x)
$$

A related approach which does not require axiomatization of C is supplied by Reiter's default logic (Reiter, 1980; Reiter & Criscuolo, 1983) which relocates the consistency check to special-purpose inference rules that supplement the usual first order set. The rule in (1.7) is glossed, "if it is known for some x that it is a bird, and if it is consistent to assume that flies, then it can be concluded that it flies."

$$
(1.7)\qquad \frac{\mathit{bird}(x)\colon \mathit{flier}(x)}{\mathit{flier}(x)}
$$

Default logic is a quite general framework, but is also undecidable.

The approach to dealing with defeasible inference that is most compatible with the systems that are the central focus of this thesis is conditional logic (Boutilier, 1989; Delgrande, 1988, 1990; Cavedon, 1995). Conditional logics augment classical systems with an additional implication relation which essentially means "typically". This approach will be described in more detail in Chapter Two of this thesis, but in quick gloss, advantage is taken of a conditional operator in addition to material implication, such that (1.8) means "birds typically fly." In terms of the possible worlds semantics underneath, it actually means "In the most normal worlds, if x is a bird then it flies."

$$
(1.8)\qquad \mathit{bird}(x) \Rightarrow\ \mathit{flier}(x)
$$

The reason for compatibility is just the fact that 'typicality' is made a primitive in the representation language. This is essentially what default inheritance reasoning provides as well, but inheritance reasoning is in a propositional setting. Another family of approaches to default reasoning are based on statistical interpretations of typicality (Bacchus, 1989; Pearl, 1987, 1988); however, I omit those from consideration here.

1.2 Inheritance Reasoning

Default path-based inheritance reasoning is basically a propositional family of logics whose syntax is usually given in graphical form. Typicality is represented in the language with default links as in (1.9).

(1.9) $bird \longrightarrow flier$

Links connect both types and individuals (1.10). Chains of links are called *paths*. Individuals can occur only as the first node of a path. Links may be positive or negative, and a negative link can occur only as the last link in a path (a negative path is one whose last link is negative).

(1.10) $Jonathon \longrightarrow bird \longrightarrow flier$

The essential terminology associated with inheritance reasoning will be presented in Chapter Two and formalized in Chapter Four of this thesis. Conclusions supported by an acyclic network composed of inheritance links correspond to paths permitted within the network. An advantage of inheritance reasoning is that inference can be polynomial (Horty, Thomason, & Touretzky, 1990).[1] The specific family of reasoners considered in this thesis are those defined by (Touretzky, 1986; Touretzky, Horty, & Thomason, 1987; Horty & Thomason, 1988; Horty et al., 1990).

Proof theoretic issues ensue from the fact that sets of defaults encoded in a network can contain conflicting paths, as the example used so far already suggests (see Figure 1.1 for the corresponding inheritance network). Because inheritance

[1]Niemelä and Rintanen (1994) present results demonstrating that certain restrictions on the ordering of information in knowledge bases can offer tractability to systems like default logic as well.

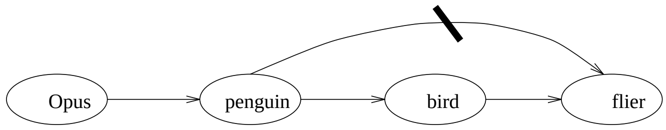


Figure 1.1: Opus Doesn't Fly

networks are directed acyclic graphs, they do naturally encode hierarchical information which can be used to arbitrate among conflicting paths, the 'most specific' path taking priority when such a judgement can be made. An obvious metric of specificity is path length, hence the definition of shortest-path reasoners. However, Touretzky (1986) noticed that shortest path reasoners are not stable in reaching the same conclusions about some networks and the same networks with links added corresponding to conclusions that were available in the original.

Definition 1 (stability) *If* ? *is an inheritance network and* $\vdash$ *is the permission relation for some reasoner* $\mathcal{R}$*,* $\mathcal{R}$ *is stable iff when* ? $\vdash \pi$ *and* ? $\vdash \pi'$ *then* ? $\cup \{\lambda(\pi)\} \vdash \pi'$*, where* λ *is just the link corresponding to the conclusion sanctioned by* π*.*

So, for example, since the network depicted in Figure 1.1 permits a path from *Opus* to *bird*, it should be possible to add a corresponding link to the network as in Figure 1.2. If a reasoner operating over the network is stable, then the

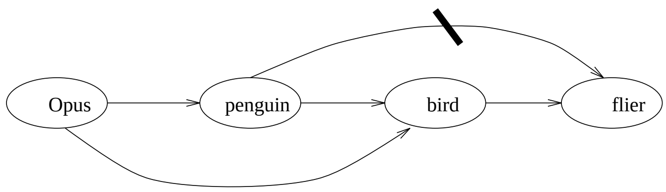


Figure 1.2: Opus Still Doesn't Fly, Or Maybe?

same conclusions will be reached. Clearly, a shortest path reasoner is not stable because in the second network it can no longer reach an unambiguous conclusion. It is arguable (see Chapter Three of this thesis, for instance, for an argument based on the way people respond), that the 'redundant' link actually should be regarded as first-class information, however Touretzky (1986) was motivated to devise an *inferential distance ordering* which basically ignores the contribution of the redundant link. Given conflicting paths, the inferential distance ordering prefers for a node to inherit properties from the *closest* node, and given nodes x, y, z, x is closer to y than to z, iff there is an inferential path from x through y to z. Thus, the *penguin* node is closer to *Opus* than the *bird* node is, and *Opus* should thus unambiguously inherit the property specified through the negative link.

Subsequent work on inheritance reasoning has led to consideration of more complex problems and has given rise to a proliferation of reasoning strategies. Touretzky et al. (1987) outline a space of variations. This thesis attempts to extend and unify this body of work by providing a set of definitions parameterized within the same formal framework that make clear how the various systems relate to each other. The second chapter of this thesis outlines the variations that are of interest here in relatively informal terms, but presents enough formal material to be able to clarify some misleading claims that have been published regarding complexity results. Essentially, it is about the complexity of inference based on the direction of path construction (whether a path is defined as a path extended by a link, two overlapping paths, or as a link extended by a path — the former is known as upwards reasoning, variations of the second are known as double chaining or downwards reasoning, and the latter has not been addressed in the literature). Some of the NP-completeness results that have been applied to downwards reasoning apply only to the double-chaining variety, not to the variety of downwards reasoning that is symmetric to upwards reasoning, introduced in this thesis. Chapter Four of this thesis presents a uniform logical definition of various reasoning systems by demonstrating how they are obtained from modifications to a logic based re-definition of a basic system (the one originally defined by Horty et al. (1990)). An advantage of this presentation is the relatively clear implementation in a logic programming environment due to the textual correspondence in the definitions and code; indeed Prolog implementations are supplied for a number of the systems. Presenting a uniform framework for specifying inheritance proof theory also has implications for the articulation of a uniform semantics parameterized to the proof theory, analogous to the way that accessibility relations in the possible worlds semantics are parameterized to the presence of modal axioms in the corresponding logic. In fact, Chapter Five of this thesis aims to present such a semantics for at least a subset of the systems considered in Chapters Two and Four. To my knowledge, this is the first unified semantics for a family of inheritance reasoners. Some specific inheritance reasoners have been provided with semantics, but not in a general enough way to make it clear how proof theoretic variations correspond to semantic ones. While this work does not exhaustively characterize reasoning in the entire space of possible inheritance systems, it does provide a substantial start.

1.3 Information Channels

The model is constructed using channel-theoretic tools provided by researchers in situation theory to account for the behavior of natural constraints as well as to model different general frameworks for representing information flow (Barwise, 1989, 1991, 1993; Barwise & Seligman, 1994). In the original formulation from situation theory (Barwise, 1989), constraints were used to interpret conditionals like those in (1.11) and (1.12) which clearly are related to generics.

(1.11) Snow means the sidewalks are slippery

(1.12) If it is snowing, then the sidewalks are slippery

Barwise interprets (1.11) with the constraint (1.13).[2]

(1.13)
$\forall l$ where l is a location parameter and s refers to a situation, $S' \Rightarrow S''$
where $S' = [\mathbf{s}' \mid \text{in } \mathbf{s}'\text{: at } \mathbf{l}\text{: snowing; 1}]$ and
$S'' = [\mathbf{s}'' \mid \text{in } \mathbf{s}''\text{: at } \mathbf{l}\text{: slippery sidewalks; 1}]$

Essentially, then, applying the constraint in (1.13) to a space-time gives the interpretation of (1.11). In this case, it is entailed that $\mathbf{s}'$ is part of $\mathbf{s}''$, but in general terms constraints convey information about the situations which support the types in the antecedent and consequents of the constraints.

Channel theory provides more structure to the analysis. Channels are tokens which connect situations and through which information flows from one to the

[2] $S = [\mathbf{s} \mid \text{in } \mathbf{s}\text{: at } \mathbf{l}\text{: slippery sidewalks; 1}]$ means that S is the type of a situation in which the sidewalks are slippery. The $\mathbf{s}$ is basically the variable that is lambda-abstracted.

other. A less realist conception of a channel is just that it is the reification of the binary relation between situations that are related by a constraint. The essential feature, though, is that channel types are objects which support constraints as their types such that all tokens of a channel type support at least all the constraints that the channel type does. Clearly this is not a reductionist account of natural constraints, but a realist one. Natural regularities are modeled by channels. Thus, if snow means that sidewalks are slippery it is because there is an information channel of a particular sort that connects snow situations with slippery ones. That is, if there is a constraint which can be used to indicate information about situations it is because that constraint is the type of a real connection between those situations. The reification of the connection is the central feature of channel theory that is important to this thesis because it is the basis of the semantics provided. This feature is what enables the convenient account of defeasibility: an informational constraint like $bird \Rightarrow flier$ can be the type of a channel which connects tokens, but it needn't require that the connected tokens are actually of the indicated types. Channel theory provides a way of allowing reasonable specification of conditions under which a constraint can be said to hold without requiring that it be correct on all tokens that might be classified in a particular way as a result. In this regard the framework has some similarity to the intensional semantics that Schubert and Pelletier (1987) offer for generics, which also makes a bit more flexible what it is to be in the extension of a predicate, although they manage without reifying connections (see also Cavedon & Glasbey, 1994).

Inheritance links are modeled by constraints which are the type of information channels which meet particular conditions (which are spelled out in Chapter Five of this thesis). Inheritance networks are translated into structures comprised of sites, types, and channels, as well as relations among these entities. In particular, sites can be understood as situations, tokens of the situation types given as the types. An indicating relation, like a default statement, can hold in general of its situation types while failing to hold for specific situation tokens. Situation tokens are related to each other through channels to create signaling relations. Information flows from one situation to another through a channel of the right type. Definitions are provided which model various sorts of path conflict in the inheritance reasoner as different conditions of information flow in the channel-theoretic model. Paths permitted by particular theories are interpreted by appropriately constructed composite channels in the corresponding channel-theoretic model. It is shown, for instance, that a path is permitted by the reasoner of Horty et al. (1990) if and only if there is an informative channel type corresponding to the link implicit in that path in the interpretation of the network when closed under the corresponding composition on channels. That is, a link is interpreted by a channel, and a path is interpreted by a composite channel. Constraints on path permission are interpreted by constraints on channel composition. Chapter Five details all this in relation to a subset of the reasoning variations considered in Chapters Two and Four. The advantage of using channels is that they supply exactly the right level of modeling to interpret a large family of reasoners within a uniform framework.

1.4 Motivations

Defaults are an important representational device that allow concise representations in formal logical frameworks. They are important because concision abets rapid prototyping of systems founded upon them, because they can offer accompanying processing gains to implemented systems and because they offer a more direct model than classical systems of human reasoning. Default logics and default reasoning have been incorporated into a number of current theories and computer implementations based on them. Paradigmatic morphology (Calder, 1989, 1990) uses defaults in lexical specification of morphology. Lascarides, Briscoe, Asher, and Copestake (1995) and Lascarides and Copestake (1995) present implemented theories of order-independent default unification for typed feature structures which helps pave the way for natural language processing systems based on Head-Driven Phrase Structure Grammar (HPSG, Pollard & Sag, 1987, 1994) to incorporate defaults. Logics of defaults have also been applied to other problems of linguistic representation such as temporal coherence in discourse (Lascarides & Oberlander, 1993). Defaults are only a small part of a large literature on nonmonotonic reasoning, and in fact have any number of practical and theoretical applications (for a review, see Pelletier & Elio, 1993). Inheritance in particular is also utilized extensively. Inheritance in natural language processing has recently been the topic of a special issue of *Computational Linguistics*, as well (for an extensive survey of applications, see: Daelemans, Smedt, & Gazdar, 1992). It is an indispensable tool for concise lexical specification; it is utilized in the lexical specification languages provided by both Paradigmatic Morphology (Calder, 1989, 1990) and DATR (Evans & Gazdar, 1989), and in organizing the lexical hierarchies of Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag, 1987), as well as in specifying the feature theory which Pollard and Sag (1994) rely upon (Carpenter, 1992). Both Calder's system and DATR admit default as well as monotonic inheritance with mechanisms for sorting out the inheritance of conflicting information. It is important to fully understand the mechanisms that drive inheritance in systems like DATR. A lucid understanding of inheritance clearly has practical and theoretical implications for a range of topics in computational linguistics. Default logics make it possible to reach conclusions based on normative facts in the absence of contradictory information, and although default logics nearly always sanction conclusions which are invalid by the standards of classical two-valued logics, these conclusions are what people consider *reasonable* inferences. Thus, default logics also supply a formal model of human reasoning that is a better approximation than classical logics can supply. Contrary to the claim of Chater and Oaksford (1990), who take Reiter's Default Logic (see Reiter & Criscuolo, 1983) as coextensive with default logics, this is bolstered by sometimes efficient decision procedures associated with constructing inferences that can be polynomial (Horty et al., 1990) and in some cases linear (Niemelä & Rintanen, 1994) (note that Lascarides and Copestake (1995) have a polynomial default unification). With efficient decision procedures, default systems have a strong claim as cognitive models since human reasoning is generally rather efficient.

While default logics do approximate human reasoning better than classical systems are able to, subsequent work in the details of the logics seems to ignore this fact. Elio and Pelletier (1993, pg. 407) point out:

> *Despite the acknowledgement by the artificial intelligence community that the goal of developing non-monotonic systems owes its justification to the success that ordinary people have in dealing with default reasoning, there has been no investigation into what sorts of default reasoning ordinary people in fact employ.* Instead, artificial intelligence researchers rely on their introspective abilities to determine whether

or not their system ought to embody such-and-so inference.

Elio and Pelletier (1993) present the first experiments into evaluating the predictions of the AI literature with respect to human reasoning with default information, specifically the prediction that given a default theory and an object which is an exception to the default, the exceptional object will have no influence on the applicability of the default to other objects. Elio and Pelletier (1993) also present the first pilot experiments testing inheritance reasoning with respect to the same prediction. Hewson and Vogel (1994) presented the first comprehensive evaluative study of default inheritance reasoning, considering its psychological plausibility as a model of human reasoning with sets of defaults. Chapter Three of this thesis presents the results of Hewson and Vogel's (1994) experiment as well as of an additional experiment which used the same materials but on a wider range of subjects. These experiments report mixed results in terms of supporting particular extant inheritance proof theories, but some of its principles are supported. While there does exist an argument that because inheritance and other default reasoning is invalid in classical terms, it is faulty, and that it therefore is not important to provide logics that implement such reasoning, there is also the reply that if not valid, such reasoning is reasonable and effective and is therefore worth the effort of studying.

1.5 Outline of the Thesis

Working with the assumption that some sort of logic other than a classical two valued system is required for a proper semantics for natural language generics, as well as to represent human reasoning with generics, Chapter Two gives a fairly comprehensive introduction to the class of logics of interest to the thesis: path based default inheritance reasoning. It outlines some of the foundational assumptions of the approach, and carves out a space of reasoners that expands the set considered by Touretzky et al. (1987). The chapter is as informal as is possible (though some definitions about path construction and permission are essential) to convey the intuitions about inference on inheritance networks while still being able to clarify some of the complexity results associated with default inheritance reasoning. While this is a contribution of the thesis, it is after all just a clarification. However, a later chapter uses the observation that negative complexity results associated with a class of downwards reasoners are actually tied only to a subclass of them, and introduces a new downwards reasoner that is symmetric in path chaining strategy to the polynomial system of Horty et al. (1990). Chapter Two also outlines the main approaches to providing a semantics to inheritance reasoning: update semantics (Veltman, 1991, 1994) (to keep the discussion in this chapter at a fairly uniform level, an appendix to the thesis gives a preçis of the formal apparatus Veltman (1994) uses), conditional logics (Delgrande, 1988, 1990), and conditional logics with preference relations on models (Boutilier, 1989).

Chapter Three addresses the issue of psychological plausibility. First a couple of informal experiments are presented to compare the relative success of classical and default interpretations of implication in predicting human behaviors, based on responses to *modus tollens*. It seems clear that the tendency to 'incorrectness' associated with misapplications of *modus tollens* is consistent with a defeasible interpretation for implication. In fact, this seems to be true as well for misapplications of *modus ponens* in which people reason from *If P then Q* and *P is true* to the conclusion that Q is only sometimes true. The chapter presents the work of Hewson and Vogel (1994) and analyzes further proof theoretic issues that were represented in their materials and data but which they did not report. A second experiment is also presented using the same materials on a larger (and all-adult) subject pool. The chapter identifies foundational assumptions of inheritance reasoning that are supported, and others that are not. This makes it difficult to adjudicate among conflicting proof theories, as was part of the original intention. Some alternative plausible proof strategies are also suggested.

The fourth chapter of the thesis re-presents the proof theoretic alternatives supplied by the family of reasoners considered in Chapter Two and Chapter Three in a more formal setting by providing a parametric definition for inheritance into which different proof theoretic alternatives can be realized. The chapter shows what modifications are required in order to achieve certain reasoning strategies, and also uses the uniform framework to initiate a closer study of exactly how the reasoners relate to each other. Corresponding implementations are also provided for some of the reasoners. While the space of potential reasoners has been discussed in the literature before (Touretzky et al., 1987) they have not, to my knowledge, been interdefined in a uniform framework such as they have been here. Chapter Five takes advantage of this parameterization in supplying a semantics for the family of inheritance reasoners. That chapter first presents the formal modeling tools supplied by channel theory as described in §1.3 above, and provides an initial application of the tools to feature theory. The chapter proceeds to develop the tools required specifically for a model of inheritance reasoners. Essentially, inheritance graphs are interpreted by models with atomic channels corresponding to inheritance links. Various channel composition operators are introduced and in the closure of an atomic model under the appropriate composition operator there is a channel for each path or link permitted corresponding reasoner, whose permission relation the composition operator interprets. The machinery is utilized to give a semantics to the family of inheritance reasoners discussed.

The sixth and final chapter summarizes the results and discusses ways of extending them. In each of the main interests of the thesis—psychological plausibility, proof theory, and parameterized semantics—there is room to cover more systems than have been by applying the same techniques. It is emphasized that more attention should be given to conducting empirical psychological research in the area to better inform proof theory of the requirements of a psychologically plausible model of human reasoning with generics.

Chapter 2

Inheritance Reasoning

2.1 Introduction: A Psychologically Plausible Model of Inference

This chapter introduces the basic notation and proof theoretic considerations at stake for inheritance reasoning. Default inheritance has been proposed as a (sometimes; see below) tractable alternative to first order logic as a model of human reasoning with natural language generics. Brachman's (1985) arguments about the impossibility of representing analytic truths in a purely default framework noted, there is reason to think that nearly all knowledge is expressed in sentences with respect to which exceptions exist. Inheritance logics make typicality a primitive and therefore offer structurally better approximations for the semantics of sentences like "Birds fly," than do the qualified sentences in first order logic. Because of the inherent representational complexity of recording exception lists, inheritance is a more psychologically plausible model. Note that using a statistical model of typicality ("Birds fly" = "Most birds fly"), chaining generics is invalid.[1] However, people do seem to make inferences from sets of generics. Therefore, it is a task of the inheritance literature to identify inheritance logics that best capture the patterns of human reasoning with generics.[2] Chapter Three explores the psychological

[1]If most As are Bs and most Bs are Cs, it is not necessarily true that most As are Cs.

[2]Of course, this is not the only task because if human reasoning is provably incorrect in some cases, then it would be nice to develop tools that help them to reason correctly or at least more correctly.

plausibility of the model in greater detail.

Inheritance networks are often based on relationships like *is-a* and *is-not-a* among individuals and classes and are presented quite simply as directed (usually) acyclic graphs of explicit links. The definitions which specify the links that are implicit in a network are considerably more complex because the network is not tree-structured. A single node may have multiple superordinate and subordinate nodes; inheritance from each node must be accounted for, as well as for the possibility that negative links (which represent *is-not-a* relations) admit conflicting inheritance information. A plethora of systems described in the literature offer a wide variety of ways to resolve conflicting chains of explicit links in determining inheritance properties. Nonetheless their procedural definition for resolving conflicting path information clouds the relationship between styles of conflict resolution and corresponding definitions of inheritance.

Chapter Four will formalize various proof theories and relate those systems on the basis of the specifications which yield the proof theoretic differences. I follow the approach of Horty et al. (1990) and the family of reasoners they discuss elsewhere (Touretzky et al., 1987). Chapter Five relates the same systems semantically using channel theoretic notions.

2.2 Notation

Path-based inheritance reasoners provide a non-monotonic propositional system for default inference which is applicable for the analysis of human reasoning about generics like, "Birds fly." In these systems, generics are encoded as links in directed acyclic graphs, the nodes of which represent individuals, properties, or classes, and the links of which represent statements of positive or negative defaults. So, for example, Figure 2.1 depicts a default inheritance network. Let the nodes of the graph labeled A, B and C represent penguins, birds and fliers, respectively. Thus, the network represents that birds fly, that penguins are birds, and that penguins do not fly.

Inheritance has been invoked as a psychologically plausible logic for capturing human reasoning with generic information. Arguably, there are no contentful assertions that are strictly true under universal quantification. Instead, such truths

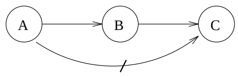


Figure 2.1: A Simple Inheritance Network

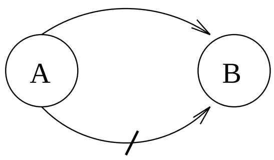


Figure 2.2: An Inconsistent Network

can be seen as typicalities, statements that hold in spite of exceptions. One approach (as in conditional logic approaches) is to analyze generics as universal implications that hold as material implications in the correctly restricted set of cases (thus eliminating exceptions). The approach of inheritance reasoning, in contrast, is to take typicality as a primitive, and to develop a nonclassical propositional logic of reasoning with such statements. The approach is motivated by the ubiquity of hierarchical arrangements of information in popular and scientific organization of knowledge. However, in part because the primitive is typicality and not strict implication, the approach leads to a nonclassical system. One feature of inheritance, is for instance, the localization of inconsistent information. Networks like the one in Figure 2.1 contain contradictions when classically interpreted, but using typicality as the primitive, it is possible to reach an unproblematic conclusion that penguins don't fly. Even if there were directly conflicting information, such as in Figure 2.2, this need not propagate as in classical systems to warrant any conclusion at all.[3]

(2.1) Linguists use handouts.

(2.2) Computer scientists don't use handouts.

[3]In fact, it seems that if there is ever going to be a general semantics for natural language, it will have to be in a framework that can represent local inconsistencies. No one would argue that Maugham's (1919) *The Moon and Sixpence* is meaningless because it contains the sentence, "There is no object more deserving of pity than the married bachelor" (p. 163), quite the contrary!

(2.3) Computational linguists use handouts and overheads.

(2.4) Computational linguists don't use overheads.

(2.5) Computational linguists are computer scientists and linguists.

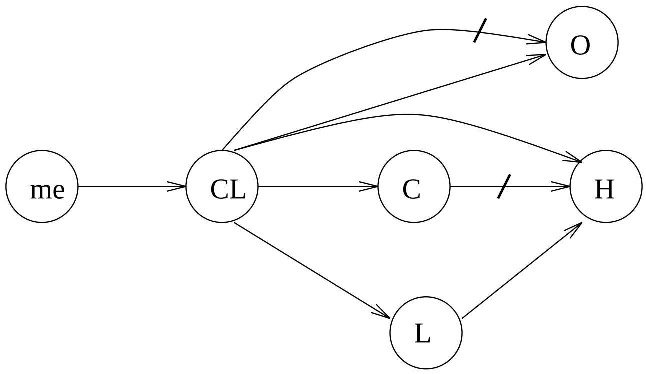


Figure 2.3: Network Representation

For example, consider the natural language generics in (2.1–2.5)(which, note, are mutually inconsistent in classical terms, and certainly typicalities rather than universal truths), represented in an inheritance network shown in Figure 2.3. Any number of inheritance reasoners will reason about these to conclude about me that it's ambiguous whether I use or don't use overheads, but that I typically use handouts. However, it is not possible to conclude arbitrary propositions in any such system. The localization of inconsistency, a form of *paraconsistency*, is deemed to be a more psychological plausible notion than the classical treatment of inconsistency. Additionally, the attractive computational properties of some of these systems (polynomial time decision) even with greater expressivity than classical propositional logics[4] make inheritance reasoning a serious candidate for consideration as a psychological model of reasoning (note that the various nonmonotonic

[4]This requires some qualification, as inheritance systems typically lack disjunction and general negation. However, since inheritance does allow typicality as a primitive and since individuals can be represented in the language, inheritance has aspects that make it more expressive than the propositional calculus. On the other hand, ignoring the interpretation of links as 'typically', inheritance can be seen as a subset of the monadic predicate calculus. In this light, the efficiency of inheritance can be seen to follow from the prohibition on cycles, and other topological features of allowable argument structures (see Niemelä & Rintanen, 1994).

logics (McDermott & Doyle, 1980) and default logics (Reiter & Criscuolo, 1983) are undecidable).

Various arguments about the 'correctness' of proof theoretic strategies are made in the literature by appeal to introspective judgements of plausibility. Chapter Three of this thesis will explore some of these issues with some rigor through experiments involving more than logician's introspections.

Inheritance reasoners define methods for reaching conclusions implicit in graph representations of sentences.[5] Implicit conclusions correspond to paths through the graph that are distinguished as *permitted*. As an example, an easily stated theory of inheritance is shortest path reasoning, in which the conclusion of a graph that is not simply linear is taken to be the conclusion that corresponds to the shortest linear path through the entire graph. Touretzky (1986) has shown this form of reasoning to be formally undesirable; however, most inheritance reasoners agree with shortest path reasoning in simple cases, and determine, for instance, that the potential path in Figure 2.1 from A to C through the intervening node B is not permitted because it is *preempted* by the *more specific* information represented by the direct negative link from A to C. However, when it comes to more complex graph topologies, different inheritance reasoners diverge considerably on which paths should be permitted from a given graph (Touretzky, 1986; Touretzky et al., 1987; Boutilier, 1989).

A wide range of path based reasoners exist in the literature (Sandewall, 1986; Stein, 1989; Touretzky, 1986; Horty et al., 1990; Boutilier, 1989; Geffner & Verma, 1989) and axes of variation among many of the systems have been outlined (Touretzky et al., 1987). In this thesis, the system of Horty et al. (1990) (the system they describe is refered to herein as $H90$) is taken as a basic system to work from. The axes of variations range from syntactic properties of reasoners (on-path or off-path preemption) to more complex semantically motivated differences (restricted skepticism vs. full skepticism). This chapter explains these properties in informal terms through graph topologies, setting the stage for Chapter Four to identify exactly how these properties of reasoners emerge from the reasoners' formal definitions (by

[5] In what follows, sets of default sentences are often referred to in terms of structural properties of those sets when represented in graphic notation; in these terms, theories are directed acyclic graphs.

providing a parameterized definition for path-based inheritance reasoners in which setting different parameters in a basic system yields each of the possibilities).

2.2.1 Some Formalities

Default inheritance networks are composed of *nodes* and default *links*. Nodes represent individuals, concepts, and properties, and links represent the classification of connected nodes. A link between two nodes in the form $A \longrightarrow B$ denotes the classification: As are typically Bs (As are Bs). A link with a slash through it, $\not\longrightarrow$, is a negative link. A *network* is a directed acyclic graph of positive and negative links, and is used to represent propositional knowledge of a "tangled" or "twisted" hierarchical nature. An *inheritance reasoner* is a set of definitions that determine the conclusions justified by a given network. Conclusions correspond to the links themselves or to chains of links (whose endpoints form an *implicit link*) that are *permitted paths*. Let π vary over paths; $lastnode(\pi)$ denotes the last node in path π, and $firstnode(\pi)$ denotes the first node in path π. If π is a positive path, then $a \overset{\sigma}{\rightsquigarrow} z$ denotes the implicit link between *firstnode*$(\pi) = a$ and *lastnode*$(\pi) = z$ through the possibly empty sequence of nodes σ. If π is a negative path, then $a \overset{\sigma}{\not\rightsquigarrow} z$ denotes the implicit link between *firstnode*(π) and *lastnode*(π) through the possibly empty sequence of nodes σ, such that the last link in π is a negative link. An *expansion* of a network is the set of explicit and implicit links supported by a reasoner applied to that network. An inheritance reasoner is the set of definitions that specify which paths in a network are permitted, and the permitted paths are exactly the contents of an expansion.

Individuals (as opposed to concepts or classes) may appear only as the first node of a path, but they need not necessarily occur in the path at all. Any link in a network is a path; if the link is of the form, $p \longrightarrow r$, then it is a *positive* path, and if it is of the form $p \not\longrightarrow r$, then it is a *negative* path. If a path consists of more than one link it is a compound path. Since networks are assumed to be acyclic, every path has both a first node and a last node. The *length* of a path is the number of links that it contains. The *degree* of a path is the length of the longest chain of links between its endpoints. For a given network, if π is a positive path, $lastnode(\pi) \longrightarrow r$ is a link, and r does not occur as a node in π, then $\pi \longrightarrow r$ is a positive path as well. Symmetrically, if $lastnode(\pi) \not\longrightarrow r$ is a link, and r does

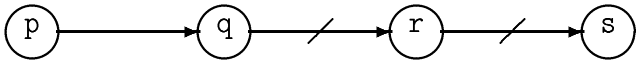


Figure 2.4: A Network with a Generalized Path of Length Three.

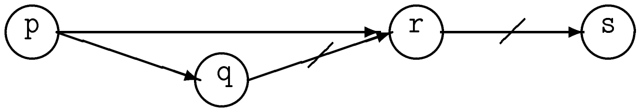


Figure 2.5: The Link $p \longrightarrow r$ Has a Degree of Two.

not occur in π, then $\pi \not\rightarrow r$ is a negative path. Since we assume π to vary over only positive paths, this means that negative links can occur only at the end of a path. The *polarity* of a path is determined by its last link: a path whose last link is negative is called a negative path, and all other paths are positive. It is convenient to refer to awkwardly long paths with strings of node labels; using this convention, "πr" and "π / r" are abbreviations for the aforementioned paths. The metanotation $\mathrel{-\!\ominus\!\!\rightarrow}$ indicates a link whose polarity does not matter and similarly for implicit links $\mathrel{\ominus\!\!\rightsquigarrow}$.[6]

Proposition 1 *If π is a path of the form $\alpha \mathrel{-\!\ominus\!\!\rightarrow} z$ having degree n, then α is a positive path with degree less than n.*

Implicit in the definition of a path's degree is the fact that there can be more than one path between any two nodes. In such cases, paths are said either to conflict or are said to be redundant to each other. If a path π is redundant with respect to another path π', then π is a subsequence of π'.

Definition 2 (Subsequences $\subseteq$) *If π and π' are two paths of the same polarity, $\pi \subseteq \pi'$ iff each node in π also occurs in π' and for any two nodes x and y such that $x \overset{\sigma}{\mathrel{\ominus\!\!\rightsquigarrow}} y$ in π, $x \overset{\sigma'}{\mathrel{\ominus\!\!\rightsquigarrow}} y$ in π'*

That is, the supersequence contains at least all of the nodes of the subsequence

[6]Note that Touretzky (1986) incorporated "don't know" links directly into his language; however, these are not used in the literature in general nor does he make extensive use of them himself. Nonetheless, it is convenient in the metatheory to have a way of talking about something that is definitely a link, even when its polarity does not matter.

and the nodes occur in the same relative order (a subpath is a special kind of subsequence in which the nodes are in the same consecutive order).

Part of the task of an inheritance system is to define procedures for adjudicating amongst these multiple conflicting and redundant paths between two nodes. Reasoners vary when settling conflict in terms of their degree of skepticism. A credulous reasoner establishes multiple expansions when faced with conflict, and in each expansion only one of the conflicting paths is sanctioned. Skeptical reasoners always resolve to a single expansion, though they vary in the number of paths that they admit to the expansion. Skeptical reasoners vary in their degree of skepticism, the more skeptical the reasoner, the smaller the size of possible expansions. The preferred paths which are chosen are said to be *permitted*. If a path is permitted in a network, then we say that the expansion of the network contains the *implicit link* between the endpoints of the path. A reasoner is *stable* if its expansions of a network are identical with the expansions of the same network with any number of its implicit links added.

2.2.2 Uniformities

Reasoners are defined to prefer (hence, permit) paths on the basis of topological considerations. An example of a topological consideration is preference for more specific information, which in network terms means being closer than other nodes to the first node in the set of paths relevant to the question. There are many axes of classification of inheritance reasoners that are interesting to focus on. The most controversial set was identified by Touretzky et al. (1987); however there are more foundational considerations than those which are interesting mainly because so few systems in the literature take exception to them. Hewson and Vogel (1994) identified a few of these in an investigation of the psychological plausibility of inheritance reasoning as a model of cognition; the findings of Hewson and Vogel (1994) will be reported and extended in Chapter Three of this thesis. This section outlines what the basic assumptions are in inheritance reasoning.

Transitivity.

This property is realized in inheritance reasoners that admit chaining of explicit links into paths corresponding to implicit conclusions. Chains of statements can be either positive or negative, and can have arbitrary (but finite) length. Not all inheritance systems admit general chaining, notably the statistically based ones (Bacchus, 1989), because it is not a statistically valid inference for defaults, although it is a practically tenable inference. Bacchus' (1989) logic mixes strict and defeasible inheritance, and limits inference through defeasible links to include only one defeasible link. Other systems also include 'certainty factors' that permit longer defeasible chains with decreasing certainty. Ballim, de Ram, and Fass (1989) present one such system which is an interesting intermediate between Bacchus' system and reasoners like H90 which admit general chaining: the system of Ballim et al. (1989) actually uses the certainty factors only to select among conflicting paths rather than setting a certainty threshold for accepting even uncontested defeasible arguments as Bacchus' (1989) logic equivalently does. It is easy to imagine other systems that incorporate a certainty calculus more flexible than Bacchus' in allowing chaining with defaults, but less flexible than Ballim et al.'s (1989) in placing a minimum threshold for certainty in chaining on even paths that lack conflicts. However, in H90 and the related family of reasoners considered in this thesis, general chaining of defeasible links is admitted. Results of Hewson and Vogel (1994) and Chapter Three confirm that people draw conclusions consistent with transitivity and suggest that there is an interesting limit on the maximum length of an acceptable transitive inference.

Acyclicity.

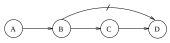
One way of stating the restriction to acyclic graphs is to stipulate that a node can occur only once in a path (thus, it is a sort of 'occurs' check). The cyclicity restriction can be slightly relaxed, as done by Geffner and Verma (1989). They release the restriction against cyclic paths and prove that in the case of negative cycles no complications are introduced which confound their definitions of path construction. A negative cycle is something like $p \longrightarrow q \longrightarrow r \not\longrightarrow p$, the cycle occurs using the single negative link allowed in the path. Since a negative link can occur only at the end of a path, a path containing a negative cycle will still have a last node. Though we can construct an endless chain of links because of the cycle, we cannot construct a *path* that has a link after the negative link. Negative cycles are less problematic than arbitrary cycles because they do not lead to paths of infinite length: all paths in a network that admits negative cycles will still have both first and last nodes. This is an interesting restriction to make on inheritance reasoning, largely because of the implications for interpretations of those systems as models of human reasoning with generics. An argument for acyclicity is that it is extremely difficult to think up taxonomies that would be best represented with cycles. In any case, whether reasoning should be restricted by acyclicity is not a property that I will investigate or elaborate in this thesis.

Negativity.

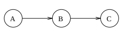
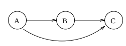
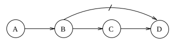
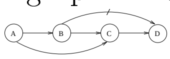
As described in Section 2.2.1, the path-based inheritance literature defines a negative path as one in which the final link is a negative link and the preceding links (if any) are all positive. This reflects the intuition that one cannot reason beyond a negative assertion of the form *As are normally not Bs*:[7] if As are normally not Bs, then As stand in no transitive relation to anything else that Bs might be. Nonetheless, general sequences that include non-final negative links can be labeled *negative chains*. Though it would be rather less well motivated than chaining of defaults, it is possible that people reason with such statements as if negativity is a *feature*, so that if As are normally not Bs and Bs are normally Cs, then As would be considered normally not Cs. Also possible in this light is that 'double negations' can cancel or intensify each other. Both potential responses are classically invalid, but negativity propagation is more pragmatically misguided than chaining: no inheritance reasoner builds in these features. Valid reasoning with negativity is one of the assumptions of inheritance reasoning that Hewson and Vogel (1994) investigated since it is interesting to know if it is a quality of human reasoning with generics.

[7] It is important to remember the difference between this and the weaker: As are not normally Bs.

Preemption.

Preemptive links are treated nearly uniformly in the literature as providing specific information that should override longer paths whose endpoints they connect. Specificity is interpreted topologically with the idea that inheritance networks encode taxonomies with most general information at the righthand side of a link, and so on throughout the hierarchy. Relative specificity is used as a way to mediate conflicting defaults. For example, , represents a network in which the inheritance literature nearly unanimously agrees to license the conclusion that As are not Ds. With this intuition, shortest path reasoning suggests itself as a simple way of computing specificity, and under the assumption that explicit links all have equal status it is an efficient way to approach inheritance. Touretzky (1986) contributed the *inferential distance ordering* to the literature in his thesis. Inferential distance is more complex in its formalization than shortest path specificity in that it can be used in systems that distinguish certain links as redundant. An example of a system in which topologically defined preemption is not incorporated is termed 'ideally skeptical reasoning' by Stein (1989). This system asumes no notion of preference in the face of conflicting arguments, even structurally defined specificity considerations.

Redundancy.

The priority of explicit links has been identified as a controversial issue with respect to stable reasoning (Boutilier, 1989). But, since Touretzky's thesis (Touretzky, 1986) it has been accepted in the literature that certain topologically identifiable links in inheritance networks are redundant since they convey no information that is not already present in longer paths, through transitivity. Essentially, certain explicit links are deemed redundant with respect to implicit links, in particular, when the explicit link expresses the same conclusion that can be drawn from a longer path. In terms of graph topology, a 'redundant' link is a direct link that connects the endpoints of another path with the same polarity. For instance, a reasoner that implements transitive inference would likely conclude from this network, , that As are normally Cs. Thus, this related network, , is deemed to contain no additional information. In a reasoner that incorporates transitivity, the conclusions implicit in both graphs are the same regardless of the information-supplying status of the redundant link; however, when graphs like these are embedded in larger networks such as, , and , a reasoner that assumes redundancy (as nearly all path-based reasoners do) will reach the same conclusions for both graphs, but reasoners that assume each direct link conveys novel information may sanction different conclusions for each graph. Touretzky (1986) argued that shortest path reasoning is unsound because of the instability that results from adding an implicit link to a network and reasoning with the result.[8] The conditional logic approach to defaults of Delgrande (1988, 1990) would find the second of the above two graphs ambiguous, and thus, Delgrande's approach is unstable (noncumulative). However, this is not a haphazard instability as in shortest path reasoning; rather, there is a strong position taken in distinguishing explicit and derived information.

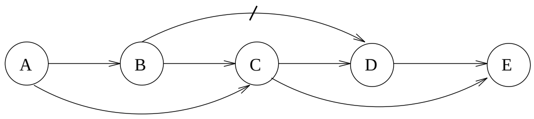


Figure 2.6: Stability/Redundancy

However, this alternative analysis of 'redundant' links is not the one that Boutilier (1989) adopts. Rather, he points out that the literature is not uniform in its topological identification of redundancy. For example, consider the graph in Figure 2.6; it is like the network just mentioned, with the exception of extension to the node E and another redundant link from $c \longrightarrow e$. Without the link $c \longrightarrow e$, H90 and Boutilier's (1989) system would conclude nothing about whether As are Es, since they do not admit chaining beyond a negative link, and specificity and redundancy for both mean that As are not Ds. However, Horty et al. (1990) define H90 so that given the network in Figure 2.6, it will conclude that As are in fact Es, while Boutilier's (1989) system reaches no conclusion for the same reason that it would without the link, under the assumption that the link $c \longrightarrow e$ is redundant. Both assume a topological definition of redundancy, but Boutilier's is simpler: for him a link is redundant if it spans a longer path of exactly the same polarity, but

[8]This property has been called the *cumulation property*: if γ is a network and l_1 is an implicit link between the endpoints of a permitted path π in γ ($\gamma \vdash \pi$ or $\gamma \vdash l_1$), then $\gamma \vdash l_2$ iff $\gamma \cup l_1 \vdash l_2$.

in H90, a link is redundant if it spans a longer path of the same polarity, and (if it is positive) no negative link contributes information about one of the nodes in the path with respect to which the link is redundant.[9] Clearly, H90 is not stable with respect to links that Boutilier would identify as redundant. Horty et al. (1990) find this an acceptable property since it gives their form of inheritance reasoning "a sensitivity to the structure of arguments that is difficult to achieve in deductive systems," while Boutilier finds it *ad hoc* and unintuitive. The semantics for inheritance reasoning given in Chapter Five is parameterized for this difference.

2.2.3 Variations

Touretzky et al. (1987) identify a number of choice points in the design of a path based inheritance reasoning system. Different decisions on these points cause different reasoners to sanction different conclusions about the same networks. The axes of variation considered herein are preemption, ambiguity, and path construction.

On-Path/Off-Path Preemption

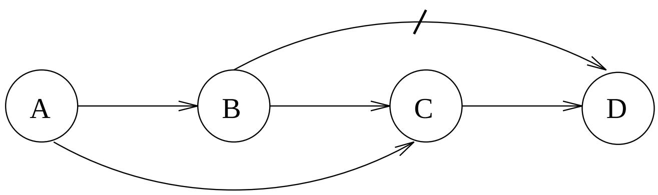


Figure 2.7: As are normally not Ds

Most inheritance reasoners include some form of preemption—a reasoner can be *on-path* or *off-path* preempting. In *on-path* preemption, the first node of a preempting link must occur on the preempted path, as in Figure 2.7 in which $b \not\rightarrow d$ is a preempting link whose initial node occurs in the preempted path *abcd*.

[9]The additional condition is not in place for the symmetric case of redundant negative links; since by the very assumption that it is a redundant negative link any positive link that would contribute conflicting information about one of the intervening nodes can at most be part of a chain that culminates in the negative link that forms the negative path that makes the other link redundant.

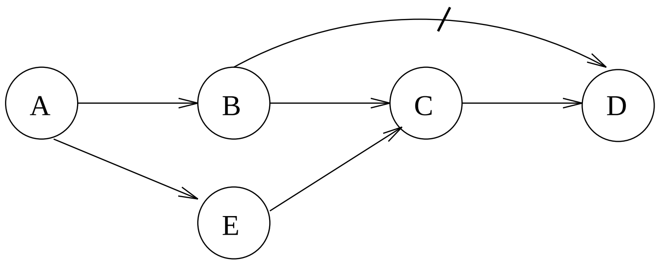


Figure 2.8: Ambiguous?

Because the path *acd* is deemed redundant with respect to the longer preempted path, it too is preempted, giving the conclusion that *As are typically not Ds.* However, for the related graph in Figure 2.8, the path *aecd* is not redundant with respect to *abcd*, and since b does not occur in *aecd*, using on-path preemption the negative path does not override the positive nor vice versa. A skeptical reasoner using on-path preemption will reach no conclusion about whether *As are Ds* or not. In a reasoner that uses off-path preemption, preemption can occur even when the preempting link does not occur on the preempted path without the preempted path being redundant. Thus, given the network shown in Figure 2.8, a reasoner that uses off-path preemption will conclude that As are not Ds.

Touretzky's (1986) original system used on-path preemption, but of course there is much debate about which is most 'intuitive'. Sandewall (1986) argues that *aecd* should be considered redundant as well, and that the link $b \not\rightarrow d$ provides the most explicit information. Touretzky et al. (1987) provide an interpretation for the graph which they claim renders the conclusions of off-path preemption questionable (a = George, b = chaplain, c = man, e = marine, d = drinks beer) since George's being a marine provides the information that he isn't a typical chaplain. Nonetheless, they adopt off-path preemption in their specification of H90 (Horty et al., 1990). This debate is reconsidered by Al-Asady and Narayanan (1993) who observe that the argument against off-path preemption given by Touretzky et al. (1987) actually runs contra to the basic principle of path-based inheritance that more specific information should override more general information: information known about George by virtue of his being a man (that men typically drink beer) is more general than the information known about him by virtue of his being a chaplain. Al-Asady and Narayanan (1993) find fault with the George-network itself, and provide an alternative hierarchy in which the same problem does not emerge. Their own Exceptional Inheritance Logic (EIL) is an inheritance system that is off-path preempting. EIL separates exceptional information from environmental information not encoded in the network, and by representing both forms in the logic, they are able to restructure networks in accordance with their intuitions. They also provide unspecified mechanisms that 'condition' (cf. Touretzky, 1986) inheritance graphs by adding nodes which locate the source of exceptional information in particular locations relative to sources of 'typical' information that conflict. This idea is close to Cripps's (1987) idea of locating exceptional information explicitly within a network at a 'minimal distance' from the source of exceptional information. Unless there is an environmental property, EIL prefers explicit exceptional information

Degree of Skepticism

Faced with ambiguity, a reasoner can either choose a conclusion nondeterministically or can refuse to draw a conclusion. A variant of the former approach is called *credulous* reasoning and variations of the latter are *skeptical* reasoners. More exactly, a credulous reasoner creates a consistent expansion for each of the possible conclusions; thus, one path is permitted in some of the expansions and the other is permitted in the rest. This thesis restricts attention to skeptical reasoners. Skeptical reasoners come in varying degrees of skepticism. A restricted skeptical reasoner reaches definite conclusions in some cases where a fully skeptical reasoner finds ambiguity. The difference is that a fully skeptical reasoner finds ambiguity in the existence of conflicting paths that have unpermitted subpaths, but a restricted skeptical reasoner requires that all subpaths of a path be permitted.

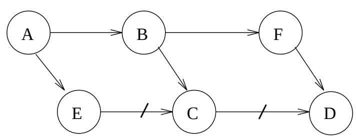


Figure 2.9: A Network with Potentially Cascaded Ambiguity

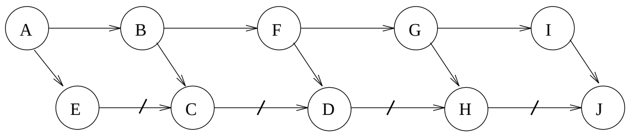


Figure 2.10: Parity Checking and Restricted Skepticism

H90 is classified as a *restricted* skeptical reasoner because it does not allow ambiguities to be *cascaded*. Consider the network depicted in Figure 2.9. A restricted skeptical reasoner determines that the ambiguity about whether a can be classified as c is sufficient to prevent the construction of a path (abc/d) which can conflict with the conclusion that a is d. A fully skeptical reasoner (or *ideally skeptical* as defined by Stein (1989)) would conclude that the conflicting paths abc/d and $abfd$ make a's d-ness ambiguous. Thus a fully skeptical reasoner has a smaller expansion than the restricted skeptical reasoner. Stein (1989) points out the curious behavior of restricted skepticism when applied to larger networks like those in Figure 2.10. As before, the restricted skeptic (H90) concludes nothing about whether a is c, but does conclude that a is d. Similarly, because a is d, H90 can conclude nothing about whether a is h, and that leaves unblocked the conclusion that a is j. Thus, ambiguity does not cascade as in the fully skeptical reasoner, but it does propagate in an alternating fashion. That the alternation has an analogy in parity checking does not render the behavior remarkably intuitive. Touretzky et al. (1987) state that it is not known whether it is possible to compute the alternative fully skeptical expansion of a network without computing the intersection of all of the credulous expansions, since this is assumed to be computationally complex in the general case.

Kautz and Selman (1991) show that the membership of a literal in all expansions can be computed in $O(n^2)$ time in the size of the network for a system that does not incorporate preemption, and this means that the intersection of expansions can be computed in $O(n^3)$ time. However, for the more interesting case of skeptical reasoning with preemption, Kautz and Selman (1991) prove that computing the membership of a literal in all expansions is co-NP-complete. This is an extremely interesting result given that Stein provides a system that also computes the intersection of credulous expansions of networks with preemption and does so in polynomial time, and that H90 (Horty et al., 1990) also computes skeptical inheritance with preemption in polynomial time. A cursory summary of NP-completeness theory (Garey & Johnson, 1979) makes it easier to appreciate the significance of these results. A problem π is in co-NP if its complement is in NP. P is closed under complementation: if π is in P so is π^c. It is not known whether P = NP, but it is generally assumed that P $\neq$ NP. However, if NP $\neq$ co-NP, then P $\neq$ NP because P is closed under complementation.[10] Thus, it seems that the elements of co-NP are 'harder' than the problems in P. Yet, problems that are co-NP-complete are easier to imagine being solvable by polynomial algorithms than those that are NP-complete since typically the co-NP-complete problems are efficiently solvable for quite large classes of instances but grind down on worst-cases. In this case, fine-grained restrictions of the problem can yield an element of P. For these results on inheritance reasoning, this suggests a closer look at the proof theoretic variants chosen by each system. The version with specificity for which Kautz and Selman's (1991) co-NP-completeness holds uses off-path preemption, while Stein's (1989) system utilizes on-path preemption (which is intuitively easier to compute), and the off-path preempting H90 maintains tractability through its restricted rather than full skepticism.

Direction of Reasoning

Unfortunately, the source of complexity is not so easily diagnosed as that. Selman and Levesque (1989) argue that the decisive complexity factor in the case of H90 is that it uses upwards reasoning, and H90 has this feature in common with Stein's (1989) system. However, here I will argue that their results are not decisive either.

Upwards reasoning corresponds to forward chaining with the point of origin being the queried node. Essentially, the definition given in Section 2.2.1 of this chapter relied on forward chaining: direct links are paths; if π is a positive path, and $lastnode(\pi) \mathrel{-\!\ominus\!\!\rightarrow} r$ is a link, and r does not occur as a node in π, then $\pi \mathrel{-\!\ominus\!\!\rightarrow} r$ is a path as well. Downwards reasoning as defined in the literature, however, relies on Touretzky's (1986) *double chaining* which can be stated as follows: direct links are paths; if $q \longrightarrow \pi$ is a path and $\pi \mathrel{-\!\ominus\!\!\rightarrow} r$ is a path then $q \longrightarrow \pi \mathrel{-\!\ominus\!\!\rightarrow} r$ is a path.

[10] Still, if NP = co-NP that does not prove that P = NP.

The difference between these modes of reasoning can be understood with respect to the differing conclusions reached about particular graphs.

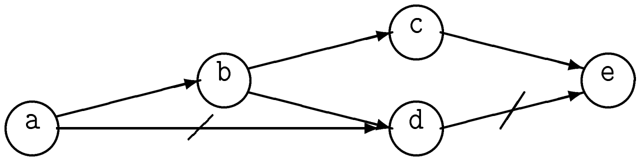


Figure 2.11: Preemption by $a \not\rightarrow d$ Permits the Conclusion $a \longrightarrow e$.

Figure 2.11 (Touretzky et al., 1987) illustrates this point. We know that in a restricted or fully skeptical upwards reasoner a will be seen as an e, since the path abd/e, which could be a path that conflicts with $abce$, cannot be formed into a path; its initial path abd is preempted. Both the restricted and fully skeptical reasoners defined above determine that a potentially conflicting path must be unpreempted. On the other hand, a downwards skeptical[11] reasoner will begin chaining at e and will stop at the ambiguity about whether b is e. There is no way to chain beyond b to notice that the leading path is preempted. If the downwards reasoner were credulous there would be an expansion in which $abce$ is held, and one in which bd/e is held, but which contains no information at all about a's e-ness since a/d preempts bd and since $a/d/e$ is not a valid path (according to the definitions given in section 2). Note that the intersection of these credulous expansions is empty. If fully skeptical reasoners are defined as the intersection of upwards and downwards credulous expansions, this could be done simply by giving the definition of a downwards skeptical reasoner. Touretzky et al. point out the differences between upwards and downwards chaining (Touretzky et al., 1987) without providing a thorough explanation of why they exist.

As a preliminary step to understanding the difference, notice the topological similarity between the network depicted in Figure 2.11 from the last paragraph and the network shown in Figure 2.9 during the discussion of degrees of skepticism. The net of Figure 2.11 could be transformed into the net of Figure 2.9 by introducing

[11]It is perhaps slightly confusing to refer to generic skeptical reasoners without indicating their degrees of skepticism, but for the present purposes their degrees of skepticism are not relevant. The distinction made is between skeptical reasoners, which for arbitrary degrees of skepticism produce single expansions when faced with ambiguities rather, and credulous reasoners, which generate multiple expansions in those same circumstances.

an intermediate node in the link a/d (modulo node renaming). Note that the restricted and fully skeptical reasoners reach the same conclusions about Figure 2.11, but different conclusions about Figure 2.9. This is because Figure 2.11 represents preemption—more specific information preempts a conflicting chain of links. Based on the difference in behavior when upwards processing these two networks, we can say the following: an upwards reasoner[12] will sometimes resolve (in cases of preemption) an apparent ambiguity at some point "higher" in the network by using more specific information from lower in the network. "Higher" is just a conception of generality included in inheritance networks—individuals, the most specific entities, are at the bottom or left, and more general classifications proceed upwards or to the right.

In contrast, a downwards reasoner will never reach definite conclusions about a network where an upwards reasoner resolves to an ambiguity. That is, since downwards reasoners begin from general and reason towards specific, it will never resolve an apparent ambiguity using the more general information already noticed. Faced with ambiguity, it will never be able to chain further towards more specific information that could resolve the ambiguity of the whole path. In schematic terms, this can be understood from the nature of the chaining specification given at the start of this section. Essentially, when using double chaining all subpaths of a permitted path will also be permitted, but that is not necessarily so with upwards reasoning. Consider a single path $\Pi = \pi \longrightarrow x = y \longrightarrow \omega$; Table 2.1 depicts the difference between upwards and downwards reasoning.

Upwards Chaining	$\underline{\quad\pi\quad}\ \underline{x}$ $\underline{y}\ \underline{\quad\omega\quad}$	$\vdash \pi x \not\supset \vdash \omega$
Downwards Chaining	$\underline{\quad\pi\quad}\ \underline{x}$ $\underline{y}\ \underline{\quad\omega\quad}$	$\vdash \pi x \supset \vdash \omega$

Table 2.1: Necessary Permission of Subpaths ($\Pi = \pi \longrightarrow x = y \longrightarrow \omega$.)

This is in fact the key difference between complexity in upwards and downwards reasoners that Selman and Levesque (1993) identify as the source of in-

[12]For the remainder of this section, "reasoner", in the context of ambiguity resolution, refers to a system that produces a single expansion—be it restricted or fully skeptical.

	skeptical		credulous	
	off-path	on-path	off-path	on-path
down	NP-hard	NP-hard	NP-hard	NP-hard
up	P	P	P	P

Table 2.2: Complexity of the various forms of defeasible inheritance. (P stands for 'doable in polynomial time').

tractability in downwards reasoners, and the reason why upwards reasoners (e.g. Stein's and H90) can be polynomial. Selman and Levesque (1989, p. 1144) summarize their results in the table reproduced in Table 2.2. However, these results are not completely decisive as they depend on the particular sort of redundancy employed in H90, discussed above. It is not clear, for example, that a system based on Boutilier's (1989) notion of redundancy would still be susceptible to the same construction that Selman and Levesque (1993) use to prove NP-hardness of downward reasoning. In fact, their construction relies crucially on the conclusions made available through a link which Boutilier would consider redundant (hence making the network ambiguous), but which Horty et al. (1990) advocate as a desirable source of instability as a nonredundant link. Further, it is difficult to reconcile Stein's (1989) claim of providing a polynomial time algorithm for ideally skeptical inheritance with specificity with the further construction of Selman and Levesque (1993) which leads them to conjecture that "any reasonable preemption strategy" will make ideally skeptical inheritance intractable. This is a stronger claim than the one reported in the last section about the complexity of ideally skeptical inheritance. The reason for pointing it out here is that the stronger claim, in conjunction with Stein's claim of actually having a polynomial algorithm to compute exactly what Selman and Levesque (1993) conjecture is not possible, indicate that there is a likely interaction of other proof theoretic features of inheritance not discriminated in the Table 2.2. I have already mentioned the treatment of redundancy as one example, but there may be others.

Chaining Complexity

One aspect of downwards reasoning that Selman and Levesque (1993) cite as a source of complexity is double chaining. When using double chaining (see the depiction in Table 2.1), all subpaths of a permitted path will also be permitted. This is in comparison to upwards reasoning, for which that is not the case. I propose to explore an obvious alternative downwards reasoning strategy that is more symmetric to the upwards. Simple downwards is defined here as the intuitive symmetric variation: direct links are paths; if π is a path, and $r \longrightarrow firstnode(\pi)$ is a link, and r does not occur as a node in π, then $r \longrightarrow \pi$ is a path as well. Hereafter, 'downwards' reasoning will be assumed to mean 'simple downwards' reasoning, and what has been traditionally called downwards reasoning will be referred to as 'double chaining'. Table 2.3 shows how downwards reasoning compares to upwards reasoning and double chaining in connection to subpaths. Clearly, this form of downwards reasoning has the same property as upwards reasoning in that all subpaths of a path need not be permitted. However, this form of downwards reasoning has not been studied before in the literature.

Upwards Chaining	π x y ω	$\vdash \pi x \not\supset \vdash \omega$
Double Chaining	π x y ω	$\vdash \pi x \supset \vdash \omega$
Downwards Chaining	π x y ω	$\vdash y\omega \not\supset \vdash \pi$

Table 2.3: Necessary Permission of Subpaths ($\Pi = \pi \longrightarrow x = y \longrightarrow \omega$.)

To understand how downwards reasoning as defined here is different, as always, it is useful to consult an example network. Consider the network depicted in Figure 2.12. H90, an upwards reasoner, will conclude that the network is ambiguous about whether As are Gs because there is no permitted path to chain the link $E \longrightarrow G$ onto (because of the ambiguous subpath from B to E). Similary for double chaining. However, using downwards reasoning, although it is not possible to construct the path $BCDEG$ due to the same ambiguity, the path $CDEG$ is available, and it is possible, particularly in a system that uses a type of redundancy closely related to H90's, to chain the link $A \longrightarrow C$ to the left end of that path, yielding an uncontested conclusion that As are Gs. H90's version of redundancy actually has this link as redundant, as discussed above, but there is a strong sense in which downwards reasoning is an inverse of upwards reasoning (almost, in fact, like reasoning 'upwards' from the most general node in the related graphs in which the links go in the opposite direction), and recall from the earlier discussion that a topologically similar link was deemed non-redundant by H90 (contra Boutilier (1989)). Chapter Four will discuss the properties of downwards reasoning in relation to the other proof theoretic variations already in the literature. Here it is important to point out just that *downwards* reasoning can have the property that Selman and Levesque (1993) identified as crucial to tractable path-based inheritance.

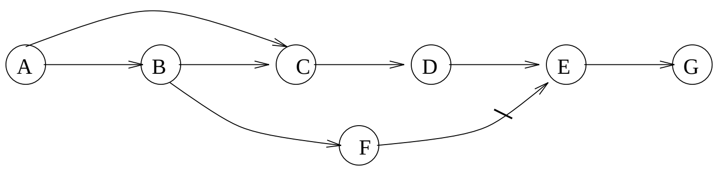


Figure 2.12: Downwards vs. Upwards vs. Double Chained Reasoning

2.2.4 The Rest of this Chapter

The purpose of this work is to explain how these variations in reasoning styles can be anchored in the formal definition of a reasoner parameterized to realize these possibilities. To this end, H90, the restricted skeptical, upwards chaining, off-path preempting reasoner of Horty et al. (1990) is adopted as standard. Chapter Four of this thesis gives the declarative specification of this system provided by Vogel, Popowich, and Cercone (1993), and also provides the parameterization that generates other reasoners and a translation of these definitions into Prolog. The remainder of the present chapter presents an outline of what I take to be the principal open problems in inheritance reasoning, the main of which is a unified semantics. This chapter will thus outline a couple of attempts in the literature to provide a semantics to path-based inheritance reasoning, but these are notable for being semantics for particular systems, and not general semantics which elucidate the related proof theories.

2.3 Discussion: Open Problems

There are quite a number of open problems in inheritance reasoning. This thesis attempts to address a few of these problems which exist in three main areas:

1. psychological investigation of inheritance as a plausible model of human reasoning with generics;
2. proof theoretic considerations that identify clear relationships among various systems so that it is clear how to, for instance, embed one within another, and particularly including complexity analysis as an evaluation metric; and
3. identification of an underlying semantic framework that unites the various proof theoretic systems that exist in the literature.

2.3.1 Plausibility of the Model

The plausibility of inheritance reasoning has been investigated hardly at all. Elio and Pelletier (1993) present results from the first pilot experiment designed to investigate how people reason about exceptional objects in the context of a larger study designed to test predictions of default logics. Hewson and Vogel (1994) present the first study that considers assumptions of inheritance reasoning directly. The next chapter of this thesis will elaborate those results on transitivity, negativity, preemption, and redundancy, and will extend those results by reporting on replications of their original experiment as well as by considering some of the more theoretically interesting features from the perspective of inheritance proof theory — features like on/off path preemption, and other specific styles of reasoning with generic information. These studies are different from the approach taken by Elio and Pelletier (1993) in that they were interested in how people reasoned about an object in light of defaults, given that it was known to be exceptional or not in particular ways. The method used by Hewson and Vogel (1994) and followed here is to consider which patterns of argument participants found decisive in general, without grounding those arguments in specific individuals.

2.3.2 Complexity Considerations

Due to the proliferation of inheritance reasoners, it is important to determine precisely how different classes of reasoners are related. Chapter Four develops a declarative specification of inheritance reasoning to mediate the "clash of intuitions" about elements in the space of possible inheritance reasoners. The relationships among various classes of reasoners are illustrated by defining the various reasoners through modifications to a basic set of definitions. This exploration allows us to identify a spectrum of skepticism in defining reasoners. Chapter Four will also provide Prolog specifications that implement the definitions of all of the reasoners considered. This Prolog specification is in close textual correspondence to the formal definitions of the reasoners, unlike the algorithmic specification of the upwards restricted skeptical reasoner of Horty, et al. (Horty et al., 1990). Textual correspondence allows each of the modified definitions to be similarly specified in Prolog, with changes to the code corresponding textually to the modifications in the definitions.

2.3.3 Semantics

The most glaring gap in the literature is in the provision of a unified semantics for inheritance reasoners. Dimopoulos (1992) claims that a series of 'process models' will do exactly that via procedural semantics but does not specify the processes nor the intended family of reasoners covered. Semantics have been offered for isolated systems–credulous or skeptical, upwards or downwards, and on-path preempting inheritance, but no attempts have been made to generalize these approaches to cover more than the individual system licensed. Interestingly, these approaches have all been developed using conditional logics as a starting point. Delgrande (1987) provided a conditional logic called N as a basis for limited reasoning with and about first order defaults. This work was extended by Delgrande (1988) to admit general chaining of defaults through making additional assumptions about the models. Delgrande (1990) provides an algorithm for constructing those models, and thereby provides a semantics to a nearly-equivalent class of inheritance reasoners (credulous and downward). Boutilier (1989) offers a conditional logic called E derived from Delgrande's N (differing in allowing nested conditionals) and extends this (Boutilier, 1993) to develop a semantics for fully stable inheritance reasoning (fully skeptical, downwards and on-path). He suggests that his approach can be modified to accommodate off-path preemption as well, though he does not specify the details, and it is not clear that such a modification would be consistent with the rest of his approach, as will be described below. Finally, Veltman (1994) is also influenced by the conditional logic interpretation of defaults, particularly Delgrande's (1990) logic, but Veltman develops this extensively into a dynamic interpretation framework called update semantics. Using update semantics, Veltman attempts to provide a semantics for the inheritance reasoner proposed by Horty et al. (1990); however, the features of Veltman's system create important differences. Veltman's (1994) system can be seen as providing a semantics for the same class of inheritance reasoners as Boutilier's (1993). The next sections give further informal detail about the mechanics of these three approaches. It probably would be possible to extend these approaches to obtain the general parameterized semantics that this thesis provides in Chapter Five, but, it would certainly require foundational changes to both Delgrande's and Boutilier's systems. Veltman's update semantics, having more complex machinery to work with could require less profound modification. However, developing a correspondence between update semantics and the channel-theoretic semantics provided here is outside the scope of this thesis. As I have said in Chapter One, I adopt the channel-theoretic approach because it provides a convenient ontology for detailing precisely and perspicuously a parameterized semantics for inheritance.

Delgrande/Boutilier — conditional logic

The main idea behind the conditional logic approach to defaults (Delgrande, 1987, 1988, 1990) is to use a modal logic with a conditional operator $\Rightarrow$, where $A \Rightarrow B$ means, "Unless there is an exceptional state of affairs, if A then B." Delgrande develops a first order modal system (which is thus able to represent strict as well as default information using the modal necessity operator for strict relations), and the propositional subset of this logic is equivalent to S4.3. The accessibility relation (E or $\geq$) associated with the logic is reflexive, transitive, and forward connected,[13] where $w_i E w_j$ denotes that w_j is 'at least as unexceptional as' w_i (w_j 'is at least as normal as' w_i). Given a sentence α, $[\![\alpha]\!]^M$ denotes the set of worlds in the model M in which α is true and that set is identified as the proposition given by α. Delgrande provides a *world selection function f* that maps from a particular world and the set of worlds given by a proposition to the set of E-accessible worlds that are at least as unexceptional (most normal). Truth ($\models$) is relative to a world and model.

Definition 3 $f(w, [\![\alpha]\!]^M) =$
$\{w_i | w \geq w_i \ \ and \ \models^M_{w_i} \alpha, \ \ and \ \forall w_j \ \ such\ that\ w_i \geq w_j \ \ and \models^M_{w_j} \alpha, w_j \geq w_i\}$

Thus the truth conditions for $A \Rightarrow B$ can be articulated as in Definition 4. This just means that $A \Rightarrow B$ is true at a world w if the worlds at least as normal as w where A is true are a subset of the worlds where B is true; that is, in the worlds more exceptional than w where A is true, B need not be true.

Definition 4 $\models^M_{w_j} A \Rightarrow B$ *iff* $f(w, [\![A]\!]^M) \subseteq [\![B]\!]^M$.

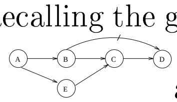
Definition 5 $\models^M_{w_j} A \supset B$ *iff if* $\models^M_{w_j} A$ *then* $\models^M_{w_j} B$.

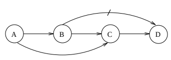
Given that the truth conditions for implication are standard (see Definition 5), it would be consistent to have for some constant c, $A(c) \Rightarrow B(c)$ and $A(c) \supset \neg B(c)$, because this just entails that there are no normal worlds where $A(c)$ is true. However, this means that the logic cannot support *modus ponens* for the conditional operator.

Since default reasoning depends on a version of *modus ponens* Delgrande (1988) offers two equivalent methods for restricting models to render *modus ponens* applicable. The first method is called the *assumption of normality*, which holds that the modeled world is among the least exceptional, and the *assumption of relevance*, in which only those sentences known to have import for the truth value of a conditional actually do affect the conditional's truth value. The efficacy of the normality method is intuitive enough. The relevance method also assumes that the world being modeled is among the most normal—among those worlds only those facts known to impinge on the truth of a conditional actually do, thus, for

[13] A relation R is forward connected iff $w_i R w_j$ and $w_i R w_k$ implies either $w_j R w_k$ or $w_k R w_j$.

example it is possible at those worlds to strengthen antecedents of conditionals like **raven** $\Rightarrow$ **flying thing** to **raven** $\wedge$ **albino** $\Rightarrow$ **flying thing** although **raven** $\wedge$ **albino** $\not\Rightarrow$ **black**. Both of these methods are defined relative to default *theories* which are modeled as tuples $\langle D, C \rangle$, where D is a set of formulae from the logic N and C is a nonempty set of first order formulae, with the intent that D encodes default possibilities about how the world might be and C represents the way the world is known to be. The normality method entails adding formulae to D, including implications corresponding to conditionals (under the assumption of normality, if $A \Rightarrow B$ then $A \supset B$); then, a sentence Δ follows by default if from the extended D it can be derived that the most normal worlds where the sentences of C are true are worlds where Δ is true. The complementary method augments C, the way the world is known to be, with contingent information from D so that sentence Δ follows by default if it is a first order consequent from the intersection of possible extensions of C (since D can lead to conflicting extensions). I have not detailed the algorithms used by either of the two reasoning methods but they work in an intuitive way by consistently expanding D or C as appropriate. Delgrande (1988) shows the methods to be equivalent in terms of the conclusions sanctioned, and notes that the algorithm given for the second method leads most easily to an implementation which he also provides (Delgrande, 1990).

Thus, the logic N augmented with the methods for generating maximal extensions to be used in making conclusions by defaults yields a clear semantics for credulous, on-path preempting, coupled inheritance reasoners, mainly because those systems make roughly similar assumptions as those built into the proof theory and semantics of N. Credulity falls out of building maximal consistent extensions—because of directly conflicting defaults there may be more than one maxima. Coupling, the agreement of subclasses on properties of superclasses, follows from Definition 4 which holds that conditionals are true just when the most normal worlds containing the subclasses are a subset of the worlds containing the superclasses. Given transitivity of $\geq$ and the assumption of normality, this entails that subclasses and superclasses will agree. In fact, the system's tight coupling is what establishes on-path preemption in addition to the stance on redundancy mentioned in §2.2.2. Recalling the graphs that distinguish a reasoner's position on both these properties, 

and 

, respectively, it is clear that

with tight coupling there is no way for the negative path from a to the rightmost node to be permitted without a and the intervening nodes on the positive chain having properties distinct from their superclasses. Thus, the system implements only on-path preemption without stability with respect to the addition of 'redundant' links. Since the desire to have stability with respect to redundant links does pervade the inheritance literature and since it is further unclear how to achieve degrees of skepticism in this framework, Delgrande's system cannot be regarded as a generalizable semantics for inheritance reasoning.

Boutilier (1989, 1993) presents an alternative conditional logic semantics for inheritance reasoning that is influenced directly by Delgrande's work but pays attention to providing a semantics to a specific class of inheritance reasoners. Boutilier uses an extension of N called variously E or $CT4D$ which admits nested conditionals. To provide a semantics for inheritance that has a more standard treatment of redundant links, Boutilier filters models with a topological definition of redundancy and uses the *minimal models* in a derived preference relation. I described Boutilier's (1993) notion of redundancy informally in §2.2.2. It is not necessary to give further formal detail here, except to mention that it is a modification of Touretzky's (1986) *inferential distance* ordering which measures the 'betweenness' of nodes to enable preference for paths through the nearest node to the node in question (see Chapter One for a definition). It is the inferential distance ordering which identifies certain links as redundant. As described earlier, Boutilier's modification just ensures that all links which have that same topology are redundant, thus achieving greater stability than is possible in the system of Horty et al. (1990) which is based on Touretzky's definition. Given a specification of a coupled, credulous, on-path preempting definition for a network ? to *support* a link λ, $CL(?)$ denotes the closure of ? with respect to the supports relation. A model of a network ? determines the truth of a positive, negative, or neutral link between any two nodes. If $\overline{\lambda}$ denotes the negation of a link through the path π in ?, it is possible to define the set of links contradicted in ? (Definition6).[14]

Definition 6 *CONTRA(?)* $= \{\overline{\lambda} : \overline{\lambda} \in$? *and* λ *is supported in* $CL(?)\}$

[14]Boutilier (1993) does not actually define this through the closure operator, using this expression instead: $\mathrm{CONTRA}(\Gamma) = \{\overline{\lambda} : \overline{\lambda} \in \Gamma$ and λ is supported in $\Gamma\}$. However, this conflates a lot of important issues: networks with their expansions and with their interpretations in models.

Because a model determines the truth of connections between all possible pairs of nodes, the articulation of a model determines a network. The preferred models of a network ? are those that satisfy the greatest amount of $CL(?)$. Thus, the preference relation on models of networks is defined to minimize the set of contradicted links:

Definition 7 *If M_1 and M_2 are models of ?, then M_1 is as* preferable *as M_2 ($M_1 \leq M_2$) iff CONTRA(M_1) $\subseteq$ CONTRA(M_2)*

A minimal model M for a network is one that satisfies ? and for which no M' exists such that $M' < M$. Nonmonotonic consequence is defined in terms of truth in all minimal models of a network.

The filter on the topological structure present in a set of sentences supported by a model is the main device for obtaining a semantics for inheritance reasoning which comes closer to any of the extant inheritance systems than Delgrande's (1990) semantics. However, the semantics provided, which does not cope with paraconisistency, prevents Boutilier from offering an exact-fit semantics for any of the main inheritance systems. Boutilier's definition of redundancy, which differs from Touretzky's (1986) as well as that of Horty et al. (1990), follows from the requirement of stability during closure under the supports relation. Without the guarantee of stability, CONTRA(?) would be impossible to define, and thus, the minimized relation would be lost. A second issue is that Boutilier's system fails to differentiate networks containing local inconsistencies. All direct-link conflicts lead to global inconsistency for the network. This is by fiat since actually $(A \Rightarrow B) \wedge (A \Rightarrow \neg B) \vdash_{CT4D} \Box\neg A$, which means that direct conflicts yield empty categories that can only vacuously combine with other arguments, unlike the behavior expected by inheritance proof theory. Modulo these caveats, Boutilier (1989, 1993) provides a semantics for stable, coupled, fully skeptical reasoning with on-path preemption. Coupling is looser than in Delgrande's system because of the structure used to filter preferred models. Skepticism follows from defining consequence in terms of truth in all minimal models of a network. On-Path preemption follows as in Delgrande's system; further, Boutilier (1993, p. 104) argues that off-path preemption is "fundamentally at odds" with redundancy.

Veltman — update semantics

Veltman's (1994) update semantics is also influenced by the conditional logic approach to defaults of Delgrande (1987, 1988). A formal preçis of the mechanics of Veltman's system is given in the appendix of this thesis. It is an interesting contrast to the other approaches named here which both determine consequence from a set of sentences taken *en masse*; instead, update semantics focuses on the sequential effects of sentences when applied to successive information states. The basic tenet of update semantics is that the meaning of sentences, defaults included, is in the effect they have on a hearer's information state. This procedural view makes the update semantics framework an excellent candidate from which to achieve a general semantics for path based inheritance, even though the formulation that Veltman (1994) gives does not provide an exact fit to any of the main inheritance systems.

The main idea is to model information states as sets of sentences 'true' or 'believed' in those states (using the power set of proposition letters, so that at the uninformed state, any describable situation is possible). The semantics of sentences is understood in terms of the sentences' effect in refining the current information state: a sentence can result in an update, which eliminates some of the possibilities (worlds); a test, which leaves the state alone but determines whether some sentence is consistent with the current state; or a 'downdate' (revision), which can reinsert previously eliminated possibilities. Klein, Moens, and Veltman (1990) and Veltman (1994) focus on the first two operations (see Lemon (1995) for an analysis of revisions). Statements of propositions and defaults eliminate possibilities from the current information state. Tests correspond to inference. The problem of stability is tackled, thusly, by distinguishing explicit information contained in a set of updates from implicit information gleaned from tests. Closure of the set of sentences under the test operation is not computed. In fact, closure is antithetical to the idea of the update semantics framework which specifies the update properties of different forms of sentences taken in sequence—given that the approach locates the meaning of a sentence in its effect on information states and that this exact effect is dependent upon the context which precedes it, any sort of closure over implicit information would yield inconsistency. The process-orientation of the framework makes it tenable as a procedural semantics for inheritance as hinted by Dimopoulos (1992). Nonetheless, Veltman takes his views on the structure of a default's effect on the state of information from Delgrande's approach to the conditional operator for defaults.

Veltman (1994) translates defaults like "birds fly" into sentences of the form $\phi \rightsquigarrow \psi$, 'if ϕ then normally ψ.' In his system, this means that the proposition ψ is a default in the domain of worlds (subsets of information states) described by ϕ, and defaults induce preferences in the information state for worlds where the propositions they express are taken to hold. Since a default can have exceptions, it is clear that this has a strong relationship with the conditional logic interpretation in which the accessibility relation gives the conditional operator the meaning, "in the least exceptional worlds where ϕ then also ψ." A more formal recount of Veltman's work is given in the appendix, but briefly I overview his approach and its relation to the others described above.

Veltman's (1994) exposition is incremental, beginning with translations of the form "normally ϕ" which are less expressive, disallowing exceptions to rules. In the more complex formulation, an information state σ is modeled as a tuple $\langle \pi, s \rangle$ relative to W, a model of the universe (the powerset of possible sentences; thus a world in W is the set of sentences held true there). The elements of the tuple are s, a nonempty subset of W, and π, a *coherent frame* (see appendix) on W. Essentially, π is a complex function that assigns a partial ordering to every subset of W. Thus, for every domain π defines the set of preferences for the domain. This subsumes the world selection function of Delgrande's system. If $[\![\phi]\!]$ is the set of worlds where ϕ is held to be true, then for any sentence ψ containing only the classical connectives and for any information state σ, the update incited by ϕ ($\sigma[\phi]$) is $-$ if $s \cap [\![\phi]\!] = \emptyset$ and otherwise the update yields a new information state with the elements of s not in $[\![\phi]\!]$ eliminated. That is, processing a non-default sentence provides more information, eliminating possible ways the world might be. In contrast, processing a sentence like $\phi \rightsquigarrow \psi$ holds s constant in resulting information states, but induces a change in the preferences given in π. Formally, this is achieved with a composition operator ($\circ$) on preference relations such that $\sigma[\phi \rightsquigarrow \psi]$ is $-$ if the domain of worlds specified by ϕ and ψ are disjoint as well as if it is incoherent (see appendix) to compose the associated preferences, and results in a refined set of preferences otherwise.

The net result is a powerful system which, like Delgrande's, puts the work of preferences for particular defaults in the semantics. The complexity of preference makes it easier to imagine a semantics tuned in this framework to the proof theories of a number of inheritance reasoners. However, also like Delgrande's and Boutilier's systems, the proof theory licensed by the most straightforward specifications as proposed by Veltman is distinct from the proof theory associated with any of the extant reasoners. In response to the question, *Does Update Semantics form a semantics for any of the extant reasoners?*, Veltman writes:

> For all inference algorithms I am acquainted with, the answer to this question is no. The algorithm for which the answer comes closest to yes is the one presented in Horty et al. (1990).

Update semantics does not provide a sound or complete semantics for that system since it licenses cascaded ambiguities as well as defeasible *modus tollens* (the former makes H90 unsound with respect to this semantics, and the latter makes H90 incomplete). Stability does not emerge as a problem in this framework because Veltman distinguishes explicit and implicit knowledge using *updates* and *tests*. Updates, as described above, create new information states with more refined information. Tests, sentences like 'Presumably ϕ', do not alter information states but rather test whether ϕ is reasonable to assert in a given information state. Thus, implicit information is never added to the current state of knowledge. While there is reason to think that as Boutilier was able to bring Delgrande's conditional logic closer to being a semantics for inheritance, further study of update semantics might also yield a semantics for inheritance reasoning (indeed, this is a topic I intend to pursue in the future), the argument of this thesis is that Channel Theory provides a more convenient set of tools for articulating a comprehensive and parameterized semantics. Since this framework has also been studied as a model of conditional reasoning (Cavedon, 1992) there is good reason to think that it can provide at least a similar handle on the phenomena. Chapter Five of this thesis will spell out the parameterized semantics in detail.

Chapter 3

Psychological Plausibility of Inheritance Reasoning

3.1 Introduction

Inheritance reasoners purport to provide a psychologically plausible model of reasoning with defaults, partially motivated by the idea that tangled hierarchies are ubiquitous in the organization of information, and partly for the descriptive power they provide in the semantic analysis of natural language generics. However, there is considerable debate in the AI literature about the "correct" definition of inheritance reasoning (Touretzky et al., 1987). Most of this discussion is based on logicians' introspective analyses of what conclusions can be drawn from any particular network of propositional default statements. Conflicting intuitions, perhaps prejudiced by interest in proof-theoretic features like computational complexity (cf. Selman & Levesque, 1989; Horty et al., 1990), are in part responsible for the lack of an accepted unifying semantics for inheritance reasoning (see Boutilier, 1989). Given the absence of a parameterized model theory, it is surprising that until very recently there have been no psychological investigations designed to elucidate the semantics of reasoning with generics with respect to the idealizations of inheritance theory. Elio and Pelletier (1993) present results about the way people classify ex-

[0]Deep thanks to Claire Hewson for collaborating on Experiment One; this chapter could not have existed without her. Thanks to Tyler Burns for connecting us with the students and teachers on SchoolNet (Lisa Callaghan and her students Christine and Kelly were quite helpful), as well as to Dianna Laurent and Norman Walker each for giving access to their undergraduates.

ceptional objects in light of default theories in relation to the way general default logics (alternative nonmonotonic systems) classify the same exceptional objects. They also present the first pilot study applying similar scrutiny to inheritance reasoners, but they do not consider other foundational claims of inheritance reasoning. This chapter presents experimental work designed to ascertain whether assumptions of inheritance reasoning about transitivity, negative reasoning, structural preemption and structural redundancy are predictive of human reasoning with generics. This same experiment was also designed to test the degree of fit of extant theories of inheritance reasoning, and those results are reported here as well. Some of the material reported here was first presented by Hewson and Vogel (1994).

3.1.1 Defining Plausibility

I take the psychological plausibility of a logic to be the degree to which it captures the reasoning patterns that people ordinarily use. Thus, I find the modeling task closer to that of theoretical linguistics in finding the right level of grammatical representation to capture linguistic phenomena as it occurs, rather than as it is prescribed. That is, just as for natural language where a prescriptive grammar is no more 'correct' than a descriptive one, neither is there a 'correct' logic of human reasoning. One can of course define logics of 'ideally rational agents', however those logics are nearly always undecidable, and therefore are perhaps even farther from the ideal for a rational agent to be driven by them. Goldman (1986) also argues that the concept of rationality should be dictated more by what human behavior demonstrates, instead of rating human reasoning as defective. I am not concerned with whether logics provide a closer morphism to the processes that actually govern human reasoning than some other formal framework (like mental models, for instance); rather, I am interested in the sort of logic that provides the best description language for expressing exactly those sentences that people are likely to agree are true or worth acting upon. This position can of course be assailed for methodological reasons by people with different goals: if the goal is to build a machine that's a 'better' or more reliable reasoner than humans are then there is little reason to develop logics with an eye on psychological plausibility. On the other hand, if the goal is to describe human reasoning in a generative formal system, then I feel this is a reasonable way of proceeding. Moreover, even if the goal is to develop a machine that reasons more correctly than humans, if it is to interact with humans it will have to have a model of human reasoning. From this perspective, plausibility can be measured in terms of the degree of fit between the conclusions licensed by a logic about a set of premises, and the conclusions reached by people. Depending upon the logic, there may be a correspondence between proof theoretic parameters and reasoning strategies that people use. There may be a partition on people's reasoning such that different sorts of logics capture the inferences of different elements of the partition, or equally possible, it may be that a person's reasoning is best described by one sort of logic in some cases and a different sort of logic in others. Because such partitioning involves a degree of consistency in people that is difficult to obtain even in experimental conditions, I believe that the within-subject multiplicity of logics is the best description. This assumption makes it reasonable to consider relative proportions of response categories between subjects as indicative of the relative proportion of within subject variation as well.

3.1.2 The Plausibility of Classical Logic

This chapter investigates the relative plausibility of various inheritance reasoners with defaults in terms of their ability to capture the inferences that people make. Chapter Two of this thesis demonstrated that the inferences licensed by the main family of reasoners I'm interested in are not classically valid. However, default inheritance does seem to provide a better model of what people do. Goldman (1986) presents research of Rips and Marcus (1977), a table summarizing the percentages of total responses for an assortment of arguments is given in Table 3.1. The percentages in each of the columns represent the corresponding proportion of participants who felt that the conclusion of the syllogism was always true, sometimes true or never true, given the truth of the first two sentences. Rips and Marcus (1977) find that people are competent with *modus ponens* (arguments 1 and 2) and its potential misapplications (3 and 4) but that they are less adept with *modus tollens* (7 and 8) and its misapplications (5 and 6). Taplin and Staudenmayer (1973) suggest that subjects sometimes interpret the conditional as a biconditional to account for behaviors in arguments like 5 and 6 (which are parallel to 1 and 2 under the $\subset$ direction of the biconditional), however this is inadequate because it does not explain the difference in the *modus tollens* arguments. Taplin and Staudenmayer's (1973) experiments presented syllogisms to subjects over twelve trials and were filtered for inconsistencies in answers over the trials; twenty-one percent of the subjects were removed in this way. Another twelve percent were consistent, but logically contradictory with respect to proposals for the truth function defining the conditional.

Argument	Always	Sometimes	Never
1. $P \supset Q$ P $\therefore Q$	$100^{\dagger}$	0	0
2. $P \supset Q$ P $\therefore \sim Q$	0	0	$100^{\dagger}$
3. $P \supset Q$ $\sim P$ $\therefore Q$	5	$79^{\dagger}$	16
4. $P \supset Q$ $\sim P$ $\therefore \sim Q$	21	$77^{\dagger}$	2
5. $P \supset Q$ Q $\therefore P$	23	$77^{\dagger}$	0
6. $P \supset Q$ Q $\therefore \sim P$	4	$82^{\dagger}$	14
7. $P \supset Q$ $\sim Q$ $\therefore P$	0	23	$77^{\dagger}$
8. $P \supset Q$ $\sim Q$ $\therefore \sim P$	$57^{\dagger}$	39	4

Table 3.1: Percent of total responses for eight types of conditional arguments († marks the answer sanctioned by classical two-valued propositional logic)

There at least two possible explanations of the lack of crispness in the tests of the other arguments besides *modus ponens*. One is that the differences can be explained in terms of people relying on more than two truth values, such that they do not feel compelled to make the expected response to the negated conclusion that they would under a two-valued system. However, the classical handling of negation in 1 and 2 slightly undermines this explanation. An alternative explanation is that they rely on a different sort of implication than the material implication, one in which it can be the case that P *implies* Q and where P can be true while Q is false. Default inheritance reasoning supplies exactly such a system: if *Ps are typically Qs* and P holds of a but Q doesn't hold of a, it just means that a is an atypical P, an exception with respect to Q. It is reasonable in such a system to use the information that *Ps are typically Qs*, and that *Q fails to hold* to conclude that *P holds*, because in the absence of information about classes like $\neg P$, P can be just one of those exceptions. Given that exceptions can exist, it may be more appropriate than speaking of 'truth' to refer to how reasonable it is to conclude something, given the knowledge known.

I tested this by partially replicating the Rips and Marcus experiment using the arguments presented in 1, 2, 7 and 8, with 4 additional variations on those arguments designed to examine reasoning with negation. In the study by Rips and Marcus (1977) subjects were asked to state whether the conclusion was always, sometimes, or never true, given the premises. Subjects each were presented with the same set of arguments (in random order) either in the form shown in Figure 3.1, or in the form shown in Figure 3.2. Both versions are equivalent to the presentation supplied by Rips and Marcus (1977) however without the symbolism for the logical constants. The difference between the two forms is in the qualification of the conclusion, subjects were asked either to state whether the conclusion was always, sometimes or never *true* given the premises, or *reasonable to conclude*. The idea is that the criterion *reasonable to conclude* need not be as strict as *true*. Thus, there should be a substantial difference in question responses between the two conditions if the non-default (classical) interpretation is the one that people ordinarily bring to the interpretation of implications and truth assessment. The results of this test are illustrated in Table 3.2. Twenty nine subjects participated via e-mail or in person. Subjects were either U.K. university students or members of the general public with access to internet (e-mail participants made themselves available via e-mail after an appeal for participation was posted to a random selection of general (non-technical) newsgroups (see Laurent & Vogel, 1994)). Subjects were arbitrarily assigned to exactly one of the two conditions. Between the two conditions (whether the subject was asked to rate whether the conclusion was true or whether it was reasonable) there is not a significant difference in distributions of answers to each problem. Examining the answer distributions within each condition, it was found that the *modus ponens* arguments elicited responses that were not likely to be random, but this is less true for the *modus tollens* arguments. The lack of significance in distributions between the two tables indicates that the null hypothesis that there is no difference in interpretations of "true" and "reasonable to conclude" cannot be rejected. However, the surprise finding is the result that even in the *modus ponens* case, people were almost as likely to respond that if P implies Q, and if P is true then Q can sometimes be false. This is in striking comparison to the results obtained by Rips and Marcus (1977) as well as by Taplin and Staudenmayer (1973) for the same problem. They found subjects much more likely to behave as predicted by material implication – Table 3.1 demonstrates complete conformance for that problem (Rips & Marcus, 1977), while Taplin and Staudenmayer (1973) had about ninety-two percent correct.

This leads me to conclude that default inheritance reasoning captures human reasoning with implications better than classical two-valued propositional logic. The observed behavior (finding Q potentially false, when P implies Q and P is true) is coherent within a default logic perspective since the P in question can simply be exceptional. There is a difference between Taplin and Staudenmayer's (1973) experiment and mine in that they tested argument classifications using "If-Then" sentences rather than "Implies" — it seems that *If A then B* is more classical than *A implies B*. I ran another version of my experiment on thirty-four undergraduate and graduate students in special education at Buffalo State College in New York using "If-Then" rather than "Implies" and did in fact obtain results closer to theirs: only nine percent gave a defeasible interpretation of *modus ponens*. However, a number of subjects changed their initial responses; considering first responses rather than final answers eighteen percent answered that *Q is sometimes true* when it is true that *If P is true then Q is true* and that *P is true*, which is the defeasible interpretation. Unfortunately, it is difficult to base analyses on patterns of answers that have been canceled. Even more interesting is the classical inconsistency exhibited in response (using final answers) to the following syllogism in comparison to the *modus ponens* case:

(1.) P implies Q
(2.) P is true
Therefore, (3.) Q is false

When asked if the third sentence is true given the truth of the first two, half of the subjects replied that (3.) is sometimes true, thirty-eight percent found it always true, and only twelve percent 'correctly' identified it as never true. Asked of the same syllogism whether (3.) is reasonable to conclude on the basis of the first two sentences a very different pattern of response emerged: eighty percent found (3.) never reasonable to conclude; fifteen percent found it sometimes reasonable and five percent found it always reasonable (each subjects received all sixteen syllogisms in this experiment — thus, the subjects were differentiating 'truth' and 'reasonable'). This strongly implies something at work other than a two-valued interpretation of truth. It is also direct evidence for defeasibility if people tend to think that Q can sometimes be false when it is true that *if P then Q* and *P is true*. Where Taplin and Staudenmayer (1973) was concerned to eliminate subjects who did not conform to the truth-functional definition of a conditional operator, I am more interested in finding a logic that captures the patterns of reasoning that are not valid within the classical two-valued framework but which people nonetheless deem reasonable. Of course, it is obvious that quantification is not propositionally formalizable, but the implication results should hold roughly for the first order case of both systems, classical two-valued systems and default systems. It remains, given the space of possible inheritance reasoners sketched in Chapter Two, to determine which proof theoretic variations of default inheritance provide the most complete coverage of the arguments that people see as reasonable.

Let P and Q be arbitrary propositions — propositions are sentences that can potentially be asserted as true or false (like *squares have four sides*). Please answer the below questions in which it does not matter what the propositions P and Q are.
Please respond to the four three-sentence problems that follow by indicating whether you think the third sentence is a true sentence all the time, some of the time, or never (check only one answer to each question).

(1.) P implies Q (2.) P is true Therefore, (3.) Q is true	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) P is true Therefore, (3.) Q is false	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) P is true Therefore, (3.) Q is neither true nor false	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) P is true Therefore, (3.) Q is either true or false	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) Q is false Therefore, (3.) P is false	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) Q is false Therefore, (3.) P is true	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) Q is false Therefore, (3.) P is neither true nor false	Always	Sometimes	Never
3 is a true sentence:			

(1.) P implies Q (2.) Q is false Therefore, (3.) P is either true or false	Always	Sometimes	Never
3 is a true sentence:			

Figure 3.1: ‘True’

Let P and Q be arbitrary propositions — propositions are sentences that can potentially be asserted as true or false (like *squares have four sides*). Please answer the below questions in which it does not matter what the propositions P and Q are.
Please respond to the four three-sentence problems that follow by indicating whether you think it is reasonable to conclude the third sentence from the first two all the time, some of the time, or never (check only one answer to each question).

(1.) P implies Q (2.) P is true Therefore, (3.) Q is true	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) P is true Therefore, (3.) Q is false	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) P is true Therefore, (3.) Q is neither true nor false	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) P is true Therefore, (3.) Q is either true or false	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) Q is false Therefore, (3.) P is false	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) Q is false Therefore, (3.) P is true	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) Q is false Therefore, (3.) P is neither true nor false	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			
(1.) P implies Q (2.) Q is false Therefore, (3.) P is either true or false	Always	Sometimes	Never
It is reasonable to conclude 3 from 1 & 2:			

Figure 3.2: ‘Reasonable’

True ("3 is a true sentence").						
Problem	Always	Sometimes	Never	Total	χ^2, d.f.=2	Significance
(1.) P implies Q (2.) P is true Therefore, (3.) Q is true	8	7	0	15	7.599	$p < .05$
(1.) P implies Q (2.) P is true Therefore, (3.) Q is false	0	6	9	15	8.4	$p < .02$
(1.) P implies Q (2.) P is true Therefore, (3.) Q is neither true nor false	1	1	13	15	19.200	$p < .01$
(1.) P implies Q (2.) P is true Therefore, (3.) Q is either true or false	12	1	2	15	14.8	$p < .01$
(1.) P implies Q (2.) Q is false Therefore, (3.) P is false	7	7	1	15	4.799	no
(1.) P implies Q (2.) Q is false Therefore, (3.) P is true	0	6	9	15	8.4	$p < .02$
(1.) P implies Q (2.) Q is false Therefore, (3.) P is neither true nor false	2	4	9	15	5.2	no
(1.) P implies Q (2.) Q is false Therefore, (3.) P is either true or false	9	3	3	15	4.799	no
Reasonable ("It is reasonable to conclude 3 from 1 & 2")						
Problem	Always	Sometimes	Never	Total	χ^2, d.f.=2	Significance
(1.) P implies Q (2.) P is true Therefore, (3.) Q is true	6	6	2	14	2.286	no
(1.) P implies Q (2.) P is true Therefore, (3.) Q is false	1	4	9	14	7	$p < .05$
(1.) P implies Q (2.) P is true Therefore, (3.) Q is neither true nor false	2	1	11	14	12.999	$p < .01$
(1.) P implies Q (2.) P is true Therefore, (3.) Q is either true or false	9	2	3	14	6.143	$p < .05$
(1.) P implies Q (2.) Q is false Therefore, (3.) P is false	5	6	3	14	0.999	no
(1.) P implies Q (2.) Q is false Therefore, (3.) P is true	2	6	6	14	2.286	no
(1.) P implies Q (2.) Q is false Therefore, (3.) P is neither true nor false	3	3	8	14	3.571	no
(1.) P implies Q (2.) Q is false Therefore, (3.) P is either true or false	8	4	2	14	4	no

Table 3.2: Ratings of Argument Validity

3.1.3 Other Investigations of Default Systems

To my knowledge Elio and Pelletier (1993) and Hewson and Vogel (1994) have been the only investigators to address the plausibility of arguments from researchers in the various default logics about the conclusions that should be reached in canonical examples. In some areas there is a consensus on the conclusions that should be reached (Lifschitz, 1989; Dorosh & Loui, 1989) although there is open debate on how complex problems should be resolved as well as controversy over simpler problems in the subset of the literature that deals with default inheritance (Touretzky et al., 1987). Pelletier and Elio (1993) give an overview of applications of nonmonotonic reasoning, and (p.21) point out:

> *Despite the acknowledgement by the artificial intelligence community that the goal of developing non-monotonic systems owes its justification to the success that ordinary people have in dealing with default reasoning, there has been no investigation into what sorts of default reasoning ordinary people in fact employ.* Instead, artificial intelligence researchers rely on their introspective abilities to determine whether or not their system ought to embody such-and-so inference. And even the 25 Benchmark Problems of Lifschitz were formulated with absolutely no regard to whether ordinary people in fact do reason in the way prescribed!

Elio and Pelletier (1993) present results of experiments in which they presented subjects with concrete instances of some of the benchmark problems. Specifically, they studied the literature prediction that the existence of an object that is an exception to a default should have no influence on the application of the default to other objects. They considered two factors which might interact with conclusions about what they call the *object-in-question* given the existence of an *exception object*. One factor is the specificity of the information that relates the exception object to the default, and the other is the degree of similarity between the two objects. Note that it is difficult to imagine approaching this investigation using abstract categories and objects; however, their between-subject variation between having the individual make a judgement and determining what would be reasonable for a robot to decide does probably approximate abstract reasoning. A sample problem (low similarity, positive form,[1] human reasoner) is given below:

[1] This means that the way in which the exception object is exceptional is specified; negative form does not detail the way the default rule is violated—in negative form, this example would not have specified where the exception object, the Craftsman drill, is.

You know	There is a Craftsman electric drill and there is also a Black & Decker electric drill.
	Electric drills are normally stored in the utility cabinet.
	The Black & Decker drill is a cordless model.
You also know	The Craftsman drill is on the workbench.

What is reasonable to decide about where the Black & Decker drill is?

The literature would predict that specificity of information known about the exception-object should not affect decisions about the object in question. Neither should the degree of similarity between the two objects. However, it was found that answers did in fact vary with specificity and similarity. Subjects favored the benchmark response more under low similarity than when the objects were highly similar. Also contrary to the predictions of the literature, subjects preferred the benchmark response more when there was low specificity about the exceptional nature of the exception-object. Elio and Pelletier (1993) also applied the same technique (reasoning about exception objects and distinct objects) in a pilot experiment about inheritance reasoning but did not report definitive findings. In one case they found ninety percent conformance to benchmark predictions and in another only fifty-three. However, the highly polarized case was one in which the prediction was a negative classification, *animals other than birds cannot fly*, and in the more divergent case it was a positive classification, *birds other than ostriches can fly.* Additionally, the problems were quite complex (ten inheritance links in the former case and twelve in the latter). There is reason to think that when problems are sufficiently complex subjects will opt for negative conclusions as a denial of the the availability of a positive conclusion (see below). In any case, I am not aware of further results in the area, apart from quite specific findings about a particular inheritance problem which will be made clear when appropriate later in the chapter. Elio and Pelletier (1994) have also studied belief revision, a related area of nonmonotonic reasoning.

3.2 Experiments

3.2.1 Overview

Two experiments are presented in parallel. The first was conducted by Hewson and Vogel (1994) and further analysis of their data is presented here. The second experiment utilizes exactly the same materials, but balances the conditions differently in order to match imbalances in the first experiment. Analyses based on pooled data have not been constructed but would be a valuable exercise with the accumulated mass of data.

3.2.2 Method

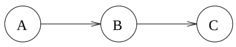
In Experiment One, seventy-two subjects were each presented with 40 problems that were designed to elicit responses which would determine whether people reason in accord with particular inheritance reasoners. The problems were comprised of sets of statements about abstract classes, as in the tests of classical logic described in the introduction. Abstract classes were used to forestall respondents from utilizing world knowledge or opinion to resolve problems on the basis of information not given in the premises. In Experiment Two, ninety-eight undergraduates were each presented with the same 40-problem questionnaire used in the first experiment. The experiment was intended to replicate the first experiment, validating the pooling of conditions in the first experiment.

Materials.

Each problem presented a set of default statements about abstract classes, followed by a question in multiple choice format; the question asked what conclusions could be drawn, based on the stated information, about the relationship between two of the classes represented (see Figure 3.3). In each problem, the respondent is asked to characterize the relationship between the 'leftmost' and 'rightmost' class mentioned in the premises. The full questionnaire contained all 40 problems, and typically took a half hour to an hour to complete.

The problems were presented either in graphical form, in sentential form, or with both graphical and sentential forms together; this created one between subjects factor—*mode of presentation*—which had three levels (graph, sentence, and graph+sentence). For each mode of presentation of a problem, the question remained the same. Thus, for the sentential presentation that the question is about the 'leftmost' and 'rightmost' classes is less obvious; however, that does not matter to the answer. Within the inheritance literature the kind of information represented in such problems tends to be presented graphically; the purpose of constructing a factor 'mode of presentation' was to determine whether the responses elicited by subjects would be affected by this factor. This has important methodological implications for investigating the psychological plausibility of inheritance reasoners. Figure 3.3 shows the graph+sentence version of problem No. 1.

$\boxed{1}$



- As are normally Bs.
- Bs are normally Cs.

What can you conclude from these statementsΓ Asterisk (*) the appropriate answer.

(a) As are normally Cs.
(b) As are normally not Cs.
(c) As are normally Cs and normally not Cs.
(d) It isn't definite whether As are normally Cs or normally not Cs.
(e) I don't know.

If you wish, explain why you reach this conclusion.

Figure 3.3: An Example Question

Subjects.

The experiments concentrated on identifying subjects untutored in logic or artificial intelligence. Thus, the initial pilot was run on individuals with long waits before their trains at Waverley Station in Edinburgh. In Experiment One, 48 Canadian students between 7th and 12th grade participated as a result of a call for subjects posted to SchoolNet, an electronic network of Canadian schools; 24 post-secondary school individuals from North America and Europe participated by responding individually to a call for subjects sent to an assortment of internet newsgroups. Experiment Two relied on undergraduate university students enrolled in English literature or composition courses at Southeastern Louisiana University. The analysis below considers the 72 subjects tested electronically in Experiment One and the 96 subjects from Experiment two.

Design and Procedure.

Experiment One. Each subject was presented with a questionnaire containing all 40 problems, randomly ordered; two random presentations (one random order and its reverse) were used and subjects were randomly assigned to receive either of these. The questionnaire contained full instructions to subjects on how to answer the problems; it was stressed that there were no right or wrong answers, and subjects were to say what *they* thought could be concluded from the information given. The mode of presentation factor created three experimental conditions: in one condition subjects received each problem in graph format, in a second condition subjects received each problem in sentence format, and in a third condition subjects received each problem in both sentence and graph formats together.

The materials were distributed electronically. As the Canadian secondary school students did not have facilities for previewing graphics, they were assigned to the sentence only condition, as were four of the other people who responded to the electronic call for participation. Text files containing the questionnaire in both orderings were emailed to the relevant teachers who randomly assigned them to students and organized their return. The graph only and mixed modes of presentation were offered to people with access to internet news, upon individual request. For each of these conditions, files containing the same orderings of the problems as in the sentence only condition and corresponding answersheets were made available via FTP in a unique location, so that no participant had access to other possible conditions in which to participate accidentally, the answersheets returned to us electronically upon completion. A general overview of possible methodologies for Internet-based experimentation is given by Laurent and Vogel (1994). Ten subjects received the graph only condition and 10 received the mixed modes of presentation. No limit was imposed on the time subjects spent on each problem. The results were pooled in the analysis; because of the different number of subjects in each mode of presentation and because of the categorical data, a nonparametric statistical test was required for the analysis, specifically, log-linear analysis.

Experiment Two. Each subject was presented with a questionnaire containing all 40 problems, randomly ordered; two random presentations (one random order and its reverse) were used and subjects were randomly assigned to receive either of these. The mode of presentation factor created two experimental conditions: in one condition subjects received each problem in graph format, in a second condition subjects received each problem in sentence format. There was an unequal number of subjects in each of the conditions because of an administrative error. Further, eight of the questionnaires were excluded from analysis because one page had been duplicated and substituted for another, thus creating a situation in which subjects would not all have had the same set of questions. The materials were distributed and returned via post. Of the ninety subjects whose materials were analyzed, seventy-six subjects were in the graph-only condition and fourteen were in the sentence only condition. The analysis utilized the same nonparametric statistical test as experiment one, log-linear analysis, which is effective when there is an unbalanced distribution of data points in the conditions. Nonetheless, in some cases the skew compromises the reliability of the resulting χ^2 analysis of the significance of differences in distributions.

3.2.3 Results and Discussion

The results were analyzed by picking out sets of problems that enabled conclusions to be drawn regarding the conformance of subjects' responses to specific predictions of inheritance reasoners. By this method it was possible to examine the extent to which people reasoned in accord with particular isolable features of these models.

Responses were coded in terms of the multiple choice answer categories (a–e, as shown in Figure 3.3), thus making it possible to directly compare subjects' responses with the predictions of inheritance reasoners. Answer 'a' always represented a positive classification of the leftmost node as typically classifiable by the rightmost node (*As are normally Bs*). Answer 'b' always represented a negative classification *As are normally not Bs* (not the negation of answer 'a', that was supplied by answer 'd'). Option 'c' classifies an assertion of definite inconsistency represented by a set of generic statements, but option 'd' expresses indeterminacy. This distinction is important to the inheritance literature since most inheritance logics classify the statements in Figure 3.4.a as inconsistent, but the ones represented by Figure 3.4.b and Figure 3.4.c are deemed inconclusive. However, both answers 'c' and 'd' can be grouped together to form a general response category meaning "not 'a' or 'b'," since the alternative option 'e', "I don't know," was also provided and exercised. In analysis of the data, categories 'c' and 'd' were collapsed into a single response category, and the category 'e' (I don't know) was excluded; thus the three response categories used were 'a', 'b', and 'c/d'. Subjects' responses are referred to as the *predicted* answer (the answer to a problem as predicted by H90 when it is 'a' or 'b'), the *complement* answer (when it is 'a' or 'b'), or an *indeterminate* answer (when it is 'c or d').

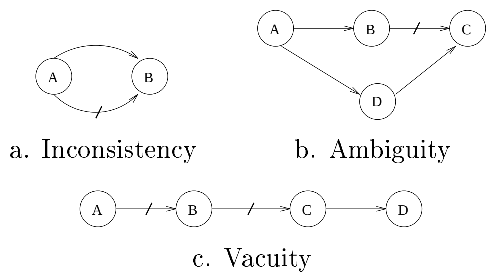


Figure 3.4: Category 'c/d': Indeterminacy

Transitivity

To test whether people reason in accord with transitivity subjects' responses to the problems with just one linear path, or a linear path with a redundant link, were compared with the responses predicted by the inheritance literature. The six problems considered were comprised of four for which the predicted answer was 'a' (positive classification), and two for which the predicted answer was 'b' (negative classification).

Experiment One.

It was found that 66% of subjects' responses conformed to the literature prediction; of the remaining 35% of responses, 11% fell into the complement category and 23% fell into the indeterminate category 'c/d'. A chi square analysis showed this difference in the proportion of responses in each category to be significant ($\chi^2(2) = 199.6, p < .01$).

These results indicate that people do tend to reason in accord with transitivity. This means that people tend to conclude from the facts that As are normally Bs and Bs are normally Cs that As are normally Cs. Having said this, there were still a remaining 35% of responses which did not accord with the transitive conclusion; this result may reflect the influence of considered statistical validity—indeed, a number of subjects indicated this as the motivation for their answers. Inheritance reasoners generally admit transitivity, although as noted in Chapter Two, some systems limit the length of chaining. For linear graphs the tendency to reason transitively fit up to the maximum length tested—three links (i.e. subjects' responses did not vary as a function of number of links). This finding supports the basic assumption of inheritance proof theory that some chaining should be admitted. More complex tests below examine the relationship between transitivity and other properties assumed by the inheritance literature.

Considering the question of whether the mode of presentation affects the way subjects respond, the proportion of responses in each category for the graph, sentence, and graph+sentence conditions were compared. Log linear analysis showed that mode of presentation did affect responses to problems for which the predicted answer was 'a' ($\chi^2(4) = 33.87, p < .01$), but not those for which the predicted answer was 'b'. For the former, the graphical conditions (graph only and graph+sentence) elicited mainly predicted responses, no complement responses, and very few indeterminate responses; however, in the sentence only condition subjects were less likely to give the predicted answer and more likely to give both the complement and indeterminate answers. Note that for each mode of presentation answers were more likely to conform to the predictions of transitivity, but the 'cleaner' conformance in the graph-only condition is stark. It suggests that the graphs lend interpretive strategies in addition to simply denoting the equivalent generics. This suggests that the natural semantics of graphs interacts with reasoning about problems when they are expressed as graphs, but problems with certain structures might lead to significantly diverging responses. This is an important point because inheritance proof theory has been developed largely with topological features of graphs in mind, and for that reason may have been misled into devices that conflict with human reasoning with generics. Although there was no significant effect of mode of presentation on responses to the problems for which the predicted answer was 'b', a similar trend was observed.

Experiment Two.

It was found that 88% of subjects' responses followed the literature prediction. The remaining 12% of responses split as follows: 5% fell into the complement category and 7% fell into category 'c/d'. A chi square analysis revealed a significant difference in the proportion of responses in each category ($\chi^2(2) = 381.225, p < .01$).

These results bolster those from the first experiment in supporting transitivity in simple problems. People tend to conclude from the facts that *As* are normally *Bs* and *Bs* are normally *Cs* (or *Bs* are normally not *Cs*) that *As* are normally *Cs* (*As* are normally not *Cs*). Note that this result is in fact stronger than the one reported in Experiment One where only 66% conformed to the literature's prediction. Loglinear analysis in that case revealed a significant effect from the mode of presentation, however that did not occur in the present experiment.

Summary. Table 3.3 summarizes the results of this section. The problem depicted there is typical of the section: the others differed solely in the numbers of links. The numbers in the columns indicate the percentages of the participants who responded to that question in each answer category. The predicted answer was positive classification (answer category 'a'): *As are typically Ds.* The caption asks a question about the graphs to make clearer what relationship subjects were looking for, but it must stressed that this is just for expository purposes here — the questions were always articulated as described in the Methods section, using a multiple choice based on a relationship about the endpoints of the set of statements (always, 'a' represented the positive classification; 'b', the negative classification;

'c/d' indeterminate classifications). That is, never did the questionnaire actually ask simply, "Are *As* typically *Ds*?" (see Figure 3.3 in §3.2.2).

Transitivity				
Problem		Indeterminate	Negative	Positive
$A \to B \to C \to D$	Exp. 1	23	11	66
	Exp. 2	8	5	87

Table 3.3: Percentages of Responses in Each Answer Category: Are *As* typically *Ds*?

Negative Paths

Negative paths can be distinguished from negative chains by defining the latter as sequences of links containing one or more negative links but which are not also negative paths. The predicted classification of negative paths is negative (answer category 'b'), and of negative chains is indeterminate (category 'c/d'). To test whether people reason in accordance with this distinction, responses to those problems with one negative path only were compared with responses to those with one negative chain only.

Experiment One.

For both chains and paths subjects gave very few 'a' responses; since 'a' represents positive classifications as in *As are typically Cs*, few 'a' responses were predicted as the problems tested contained negative links. However, whereas for negative paths there were a lot more 'b' (65%) than 'c/d' (27%) responses, for negative chains there were roughly equal numbers of 'b' (41%) and 'c/d' (47%) responses. Subjects were more likely to say 'b' and less likely to say 'c/d' for negative paths than for negative chains ($\chi^2(6) = 32.91, p < .01$), thus indicating that they do distinguish between the two. This suggests that people differentiate the validity of transitivity and general negative chaining. However, the results present some surprises: although people largely reason in accord with predictions for negative paths, the answers elicited with respect to negative chains show greater deviation from literature predictions. The equal proportion of 'c/d' (predicted) and 'b' answers to negative chains indicates a substantial tendency for people to opt for the answer predicted by set theoretic interpretations. One speculation, which provides a basis for follow-up studies, is that less abstract problems would alter this pattern of responses.

Log linear analysis revealed that mode of presentation had an effect on subjects' responses for negative chains ($\chi^2(4) = 19.36, p < .01$) but not for negative paths. For chains, in the graph only condition subjects responded 'c/d'(64%) more often than 'b'(33%), but in the sentence only condition there were equal numbers of responses in these categories (43%). (The graph/sentence condition had slightly more 'c/d'(54%) responses than 'b'(42%) responses). This result reinforces the trend observed in the preceding discussion of transitivity for responses to be more polarized in the graph conditions than in the sentence condition.

Experiment Two.

In response both to problems containing just a negative path and to those containing just a negative chain, subjects gave very few 'a' responses (3% in response to negative paths and 6% in response to negative chains), slightly more indeterminate responses ('c/d'; 11% for negative paths and 19% for negative chains), and mainly 'b' responses (86% for negative paths and 75% for negative chains), indicating a sort of 'transitive' conclusion that there is a definite negative relationship between the endpoints of the chain. This gives a different finding from Experiment One in that there is less distinction in between response patterns for negative paths and negative chains, although the 11% difference in 'b' responses is met mainly by an increase in 'c/d' responses, thus creating the same direction of the trend found in Experiment One. In fact, subjects were more likely in the case of negative chains to answer 'a' or 'c/d' than they were in the case of negative paths ($\chi^2(2) = 9.12, p < .02$). Moreover, this finding amplifies the surprise found in Experiment One: whereas in that experiment an equal amount of 'b' and 'c/d' answers were noticed, here substantially more answers were 'b' for negative chains, but by definition, negative chains are those that the literature predicts should be 'c/d'. Given the principle difference between experiments in subjects being average age, this means that the younger population from the earlier experiment peformed closer to the set-theoretic ideal of correctness. Given that the age group in Experiment Two is more representative of the adult population, this means that inheritance reasoners in general do not implement basic patterns of reasoning with negative information.

Mode of presentation was found to have had an effect on subjects' responses ($\chi^2(5) = 21.36, p < .01$). As in Experiment One, this effect existed for negative chains but not for negative paths, even though more people selected 'b' than 'c/d' in Experiment Two. For negative chains, in the graph only condition, 'b' was answered 78% while 'c/d' was given 16%; in the sentence only condition, there was a more even distribution: 'b' — 60%, 'c/d' — 37%. This result reinforces the observation that graph-only conditions yield more highly polarized responses.

Summary. Table 3.4 summarizes the results of this section. The problems there are representative of the set of problems tested in this section; the others differed solely in the numbers of links. The numbers in the columns indicate the percentages of the participants who responded to that question in each answer category. The predicted answer was negative classification in the case of negative paths *As are typically not Cs*. The literature predicted that the second set, negative chains, would be classified as indeterminate rather than with either the positive or negative classification.

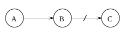
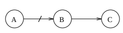
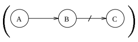
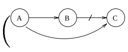
Negative Paths				
Problem		Indeterminate	Negative	Positive
	Exp. 1	27	65	8
	Exp. 2	11	86	3
Negative Chains				
Problem		Indeterminate (Pred)	Negative	Positive
	Exp. 1	47	41	12
	Exp. 2	19	75	6

Table 3.4: Percentages of Responses in Each Answer Category: Are *As* typically *Cs*?

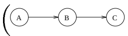
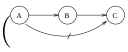
Preemptive Links

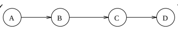
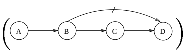
As described in Chapter Two, preemptive links are treated almost uniformly in the literature as providing more specific information that should override longer paths whose endpoints they connect. To test whether subjects dealt with preemptive links as predicted by the literature, 5 pairs of problems were compared; each comparison involved a graph and its sister graph which was identical except for the addition of one preemptive link. Thus, the task was to determine whether there was a significant difference in the patterns of responses to the two problems.

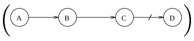
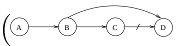
Experiment One.

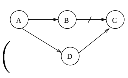
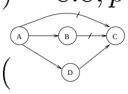
The first comparison involved the graph () and its sister (); the inheritance literature predicts that subjects should answer 'b' (negative classification) to the first of these and 'a' (positive classification) to its sister. Subjects gave mainly the predicted answer (68%) to the first graph, but there were also a fair number of indeterminate responses (27%); however, for the sister graph subjects were as likely to give an indeterminate response (42%) as the predicted response (43%). Log linear analysis showed that the difference in distribution of responses to each of these problems was significant ($\chi^2(6) = 63.31, p < .01$); thus, with the addition of the preemptive link there was a reduction of 'b' (predicted) responses which was reflected in an increase in both 'a' (predicted) and 'c/d' responses (with the increase in predicted responses being greater). This result suggests that the effect of a preemptive link is not to override the existing path, as the inheritance literature argues, but rather to add extra information which is considered along with the existing path.

Log linear analysis showed an interaction between the difference in subjects' responses to each of the problems and mode of presentation ($\chi^2(4) = 13.07, p < .01$): for the graph only and graph+sentence conditions there was a very large drop in the number of 'b' responses (from 90% and 78%, both to 0%) between the graph and its sister, accounted for by an increase in both 'a' and 'c/d' responses. However, for the sentence only condition the observed drop in 'b' responses was a lot less (61% to 20%), and was accounted for primarily by an increase in 'a' responses. Again this result confirms the observation that the distribution of subjects' responses tends to be more polarized when graphical rather than sentential information is presented.

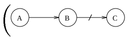
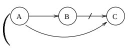
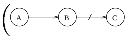
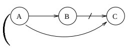
A second comparison involved the two graphs () and (). This pair is symmetric to the preceding one, and elicited an almost identical pattern of results (in terms of predicted, complement, and indeterminate responses). The inheritance literature predicts that people will answer 'a' to the first graph and 'b' to the second graph. Subjects did mostly answer as predicted to the first graph (64%), but also answered ambiguous (24%) about a third as often; for the sister graph there were roughly equal numbers of predicted (49%) and ambiguous (43%) responses. Thus between the first graph and its sister there was a decrease in the number of 'a' responses, accounted for by an increase in 'b' responses and, to a lesser extent, 'c/d' responses. As in the preceding comparison, log linear analysis showed the observed difference in the distribution of responses to each of these two graphs to be significant ($\chi^2(6) = 64.26, p < .01$). This sustains the inference that people do not treat a preemptive link as overriding an existing path in the way the inheritance literature predicts. Surprisingly, in contrast to the previous case, no significant effect of mode of presentation was found for this comparison.

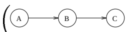
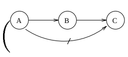
The graphs () and () were involved in the third comparison. Again, log linear analysis showed that the distribution of subjects' responses between each of these graphs was significantly different ($\chi^2(6) = 35.05, p < .01$), and this effect interacted with mode of presentation ($\chi^2(4) = 10.26, p < .05$). Subjects gave mainly the predicted answer (60%) to the first graph, though there were also about half as many indeterminate responses (31%); for the sister most responses fell into the indeterminate category (54%), and roughly half as many in each of the predicted (24%) and complement (22%) categories. In this case the addition of the preempting link revealed a decrease in 'a' (predicted) responses, accounted for by an increase in both 'b' and 'c/d' responses; however, the predicted increase in 'b' responses was smaller than the unexpected increase in 'c/d' responses. Again this result suggests that the effect of a preempting link is not to override the existing information, but to add to it. As for the effect of mode of presentation, for both the graph-involving conditions there was a large drop in the number of 'a' responses—in the graph only condition this drop was accounted for by an increase in 'c/d' responses, and in the graph/sentence condition it was accounted for by an increase in both 'c/d' and 'b' responses. In the sentence only condition the decrease in 'a' responses was less marked, and was accounted for by an increase in 'c/d' and 'b' responses. Again the graph conditions show more polarized response.

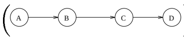
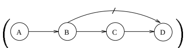
A fourth comparison examined a pair of graphs which are symmetric to the above pair, () and (). The pattern of responses elicited by these graphs was very similar to the pattern described above (in terms of predicted, complement, and indeterminate responses). For the first graph the literature prediction was answer 'b'; subjects did mostly give response 'b' (61%), but also responded 'c/d' (28%) and 'a' (10%); for the sister graph the literature predicted answer 'a', but in complete contrast subjects responded 'a' the least (only 15%), 'b' slightly more (19%) and mainly 'c/d' (66%). Thus with the addition of a preemptive link the number of 'b' responses decreased, but rather than this effect being accounted for by an increase in 'a' responses (as the literature predicts), an increase in 'c/d' responses was observed. Again the observed difference in the distribution of responses to each of these graphs was significant under a log linear analysis ($\chi^2(6) = 36.21, p < .05$). This reinforces the same conclusion, that preemptive links do not override existing links of opposite polarity, but tend to push people towards concluding there is a conflict, and choosing response 'c' or 'd'. However, unexpectedly, the effect of mode of presentation in this case was not significant though it did approach significance $\chi^2(4) = 8.8, p = .0642$.

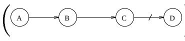
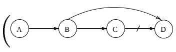
A final analysis compared graphs () and (); the distributions of responses between these two graph were significantly different ($\chi^2(5) = 22.00, p < .01$). To the first graph subjects responded mainly 'c/d' (65%) as predicted, and to a lesser extent, and each about equally, 'a'(15%) and 'b' (19%)); to its sister they responded mostly 'b' (53%) as predicted, but also 'c/d' (37%) and 'a' (10%). Thus, addition of a preemptive link in this case is reflected in a decrease in 'c/d' responses, accounted for by an increase in 'b' responses. Since this shift from 'c/d' to 'b' responses was only partial, where the literature predicts a complete shift from 'c/d' to 'b', this result also confirms the idea that addition of a preemptive link does not override the existing link. There was no effect on mode of presentation on this result. This test provides a specific point of comparison with the results of the pilot experiment run by Elio and Pelletier (1993); we found for the first of the two graphs that people mainly classified the graph as indeterminate (65%) and only 15% and 19% in each of the definite categories while in Elio and Pelletier's (1993) study of the same problem (presented with interpretations and with a different sort of question) roughly half of the people found the problem determinate, though people still split about equally between the two determinate categories.

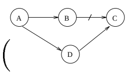
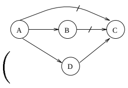
Experiment Two.

The first comparison involved the graph () and the related network (). As mentioned above, the literature predicts that subjects will respond 'b' to the first and 'a' to its sister. Subjects gave mainly the predicted answer (86%) to the first graph, but there were also some number of indeterminate responses (10%) and complement responses (4%); however, for the sister graph, the percentage of indeterminate response (54%) was much closer to the percentage of predicted response (38%). This is the same trend found in Experiment One, although there was a smaller percentage of indeterminacy in that study. Log linear analysis showed that the difference in distribution of responses to each of these problems was significant ($\chi^2(2) = 110.912, p < .01$). With the addition of the preemptive link to the first graph () creating the second graph (), there was a reduction of 'b' responses (predicted for the first) which was reflected in an increase in both 'a' (predicted for the second) and 'c/d' responses (with the increase in nonpredicted responses being greater). This result, increased indeterminacy arising from the explicit preemptive link, corroborates the finding of Experiment One that the effect of a preemptive link is not to override the existing path, as the inheritance literature argues, but rather to add extra information which is considered along with the existing path. However, contrary to the results in the first experiment, there was no effect of mode of presentation on responses. This is interesting since the effect in Experiment One was an increase in responses labeling a problem indeterminate when stated in the sentence only condition over the responses to the same question in the graph only condition. However, in the present case, indeterminacy in response to the preemptive link was the majority response under both modes of presentation.

The symmetric pair () and () elicited a similar pattern of response. If people responded the way the inheritance literature treats these graphs, they would answer 'a' to the first graph and 'b' to the second. Subjects did mostly answer as predicted to the first graph (90%), but also answered ambiguous (7%) and a very small percentage gave (3%) the complement response. To the sister graph (which had a roughly equal split in response between predicted and indeterminate classifications in Experiment One) subjects classified it mainly as indeterminate (62%), while a quarter of the people gave the predicted answer and about half as many as that gave the complement response (13%). The difference in response patterns was significant ($\chi^2(2) = 102.191, p < .01$), suggesting again that direct explicit links are not treated as preemptive in human reasoning, but somehow confounding. In comparison to the symmetric pair of problems given above, the preempted link, which was negative in this comparison, yielded more complement responses, rather fewer predicted responses and substantially more indeterminate responses (however, this difference was only approaching significance). This trend suggests a difference exists in reasoning with additional positive and negative preemptive links. There was not a significant effect of mode of presentation.

Response to the graphs () and () were quite different from the previous cases even though structurally they are quite similar. As in Experiment One, log linear analysis showed that the distribution of subjects' responses between each of these graphs was significantly different ($\chi^2(2) = 57.720, p < .01$), and this effect interacted with mode of presentation ($\chi^2(5) = 21.95, p < .01$). Subjects gave mainly the predicted answer (93%) to the first graph (cf. 60% in the first experiment), and more indeterminate classifications (6%) than complement responses (1%) (cf. 31% and 9%, respectively, in Experiment One). For the sister graph most responses were in the indeterminate category (49%) followed by the complement category (38%), with the smallest amount of people following the actual prediction (13%). By contrast, for experiment one, while roughly half of the responses were also in the indeterminate category, the remainder were evenly split between the indeterminate and complement categories. The response to the additional, preemptive, link seems to have been a substantial (but not complete) decrease in 'a' responses (predicted for the first graph) accounted for mainly by an increase in the 'c/d' responses. As in Experiment One, the predicted increase in ‘b’ responses was significantly less than expected. This reinforces evidence that additional direct links do not actually preempt the other information, but augment it in a way that perhaps leads to ambiguity and inconclusiveness. The mode of presentation effect was quite stark, though curiously not demonstrative of greater polarization of response in the graph-only condition. Rather the difference between the two graphs in the graph only condition lead to a fairly even division of responses between the indeterminate and complement categories, with a small percentage (7%) in the actual predicted category. By contrast, in the sentence only condition there was a shift from a fully predicted response (100%) to mainly indeterminacy (54%) and the rest in the predicted category (46%). There were no responses in the complement category.

The related comparison between graphs symmetric to the above pair, () and () found a very similar pattern in terms of predicted, complement, and indeterminate responses. For the first graph the literature prediction was answer ‘b’, and subjects did give response predominantly ‘b’ (85%), but also responded ‘c/d’ (13%) while giving a negligible complement response ‘a’ (2%). For the sister graph the literature predicted answer ‘a’, but few subjects responded ‘a’ (only 17%), ‘b’ (the complement response for this graph) slightly less (15%) and mainly ‘c/d’ (68%). This is virtually the same breakdown as in Experiment One. Thus with the addition of a preemptive link the number of ‘b’ responses decreased, but rather than this effect being accounted for by an increase in ‘a’ responses (as the literature predicts), an increase in ‘c/d’ responses was observed. Again the observed difference in the distribution of responses to each of these graphs was significant under a log linear analysis ($\chi^2(2) = 85.626, p < .01$). This reinforces the same conclusion, that preemptive links do not override existing links of opposite polarity, but tend to push people towards concluding there is a conflict, and choosing response ‘c’ or ‘d’. Quite curiously, there was not an effect of mode of presentation in this case. This is interesting since there was an effect for the related graphs just above, and there was similarly no effect of mode of presentation for this particular case in Experiment One either (although in that case the trend was approaching significance). It is also interesting to compare the results of the sister graph in this case with the sister graph in the preceding case: for both, the unpredicted ‘c/d’ resonse was the majority (58% and 41%), however in the case in which a negative link was added to a positive chain subjects were more likely to give the complement answer, 'a' rather than the answer predicted by preemption (32% answered 'a' and 11% answered 'b'), and in the case of the positive link added to the negative path, the answer was split roughly equally (13% and 11%). This difference was significant $\chi^2(2) = 9.213, p < .05$, and gives further evidence to an asymmetry in the handling of positive and negative information.

The final analysis of preemption compared graphs () and (). the distributions of responses between these two graph were significantly different ($\chi^2(2) = 14.253, p < .01$). The association between the presence or absence of the preemptive link and answer category was quite strong (predicting the answer on the basis of the problem structure would have reduced the error of guessing it at random by 24% (using λ_y proportional reduction in error)), although this association is not in the direction predicted by inheritance reasoners. To the first graph subjects responded overwhelmingly 'c/d' (72%) as predicted, and to a much lesser extent, 'b' (18%) and rather fewer responded with 'a'(10%). Note that this is even further from replicating the Elio and Pelletier (1993) result (half indeterminate, and an equal split between 'a' and 'b' on the remainder) than the first experiment was. To the sister graph, people responded with mainly 'c/d' (52%), slightly fewer with 'b' (43%), and some 'a' (5%). That is, with the preemptive link, subjects tended to find the graph more indeterminate than without, although still mainly indeterminate. Again, the trend towards indeterminacy is much stronger than in Experiment One where the answer distribution ran thusly, 'b' (53%), as predicted, but also 'c/d' (37%) and 'a' (10%). This finding gives further support to the idea that preemptive links supply confounding information. As in the first experiment there was no effect on mode of presentation on this result independent of the association between answer category and the presence or absence of the preemptive link.

Summary. Findings of this section are given in a different way in Table 3.5. The problems shown in the table are just those with the link that the literature predicted would be considered preemptive. The columns indicate the percentages of the participants who responded to that question by classifying it with the answer category indicated by the column heading (indeterminate = 'c/d', negative = 'b', positive = 'a'). The predicted response was positive classification for the first and fourth networks, and negative classification for the others.

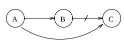
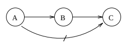
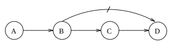
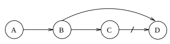
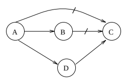
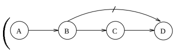
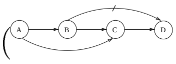
Preemption				
Problem		Indeterminate	Negative	Positive
	Exp. 1	42	15	43
	Exp. 2	54	8	38
	Exp. 1	43	49	8
	Exp. 2	62	25	13
	Exp. 1	54	24	22
	Exp. 2	49	13	38
	Exp. 1	66	19	15
	Exp. 2	68	15	17
	Exp. 1	37	53	10
	Exp. 2	52	43	5

Table 3.5: Percentages of Responses in Each Answer Category: Are *As* typically *Cs*? *Ds*?

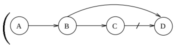
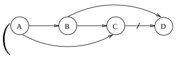
Redundant Links

The last section discussed preemption, a phenomenon in which a topologically distinguished link is supposed to override other information encoded in a set of sentences. This section considers a similar phenomenon, but one in which the literature predicts the link to be ignored.

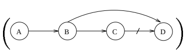
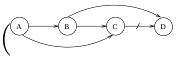
The effect of redundant links on subjects' responses was examined by comparing the responses to pairs of graphs that were identical apart from one redundant link; if subjects reason in accord with the predictions of the literature then their answers should not be affected by the addition of a redundant link. These paired comparisons could be broken into two groups based on the balance of polarity among paths through the networks—those in which the original graph had an equal number of positive or negative paths and for which the additional link would have offset the balance and those in which the original graph had only paths of one polarity or the other and the additional link created just another path of the same polarity. As expected in the latter case, it was found, for each of the paired comparisons, that subjects' responses were not significantly affected by the addition of a redundant link, nor was there an interaction of mode of presentation. The former case is more interesting to the inheritance literature since it involves comparisons between graphs like () and (), and if the inheritance literature is correct there will be no difference in response because the additional link contains no information that is not already in the original graph. An alternative proof theory in which explicit links are assumed to convey novel information might propose a method of 'path counting' in which the number of arguments in favor of one conclusion or the other determines the decision and would predict a different response between the two problems.

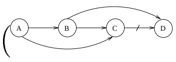
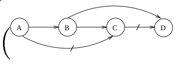
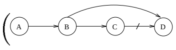
Experiment One.

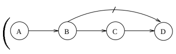
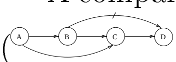
As it happens, there was not a significant difference in response, which in this case fails to support path counting. However, it cannot really be taken as evidence for the predictions of the inheritance literature because the inheritance literature predicted no change in answers between the two graphs, but further predicted a definite rather than indeterminate response to the graphs. The response patterns for the first graph were stated in the preceding section where it was pointed out that people behaved contrary to the predictions of the inheritance literature: 24% gave the response predicted by the literature ('b', negative classification) and 22% gave the complementary response, while 54% classified it as indeterminate although the literature presents strong intuitions that people will give a positive classification (answer 'b') to conclude that *As are normally not Ds.* People gave similar responses to the sister graph with the redundant link (62% indeterminate, 17% predicted and 21% complementary). Log linear analysis showed there to be no significant difference between the distributions of responses to each of these graphs. This supports the intuition presented in the inheritance literature that the additional link is in fact redundant, even though the predicted response to the graph is not borne out. To understand this, note that a 'path counting' system which incorporated a topological definition of redundancy would have predicted that both graphs be classified as indeterminate since they would have an equal number of non-redundant positive and negative paths.

Another comparison examined the pair, () and its sister () which are symmetric in polarity to the first comparison. As in that case, people answered contrary to the inheritance literature for the first graph: 66% indeterminate, 15% predicted (‘a’) and 19% complement. For the sister graph responses were: 53% indeterminate, 11% predicted, and 36% complement. Again there was no significant difference between the distributions of responses to each of these graphs. It is interesting that, though not significant, there was an observed trend in the responses to these two graphs such that with the addition of the redundant link responses shifted away from determinacy, not to the literature-predicted answer (‘a’), but to the answer that would have been predicted by a path counting method (‘b’). However, the nonsignificance of the difference gives support to the idea that the additional link does not convey novel information.

Experiment Two.

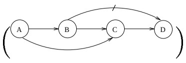
The responses to the graphs () and its sister () were compared to verify the literature’s prediction that there should be no difference in conclusions between these graphs. In Experiment One, the difference in response distributions were not significant, however in Experiment Two there was a significant difference ($\chi^2(2) = 7.669, p < .05$). In the first experiment, people found the first graph indeterminate (66%) and were roughly equally split between complement and predicted responses and found the second graph less indeterminate (53%), giving a small predicted response (11%) and a substantial complement response (36%). In Experiment Two, response to the first graph was slightly less determinate (predicted: 17%; complement: 15%; indeterminate 68%), and there was a shift in response to the graph with a ‘redundant’ link added (predicted: 8%; complement: 31%; indeterminate 61%) parallel to that found in Experiment One. In particular, with the redundant link in place, people were less likely to give either the predicted response or the indeterminate classification, and more likely to give the complement response. An interpretation of this finding which is aligned with the findings of the previous section is that redundant links are not actually treated as such. If one imagines the redundant link forming part of a second negative path between A and D, then aside from the overwhelming majority who find the graph indeterminate to start with, there is a substantial population that might be reasoning that two negative paths override the single positive path (path counting). Mode of presentation did not have a significant effect on the association between answer category and the presence or absence of the redundant link.

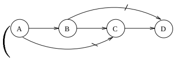
Given the earlier finding that people do not reason with negative information as predicted by the literature, that they respond paths with nonfinal negative links (and multiple negative links) as negative paths, it is interesting to compare the second graph () with a related one (), also different from the first graph () by one link, although the literature does not find the additional link redundant in this case because it contains information not already expressed in the path it spans. The response pattern to this third graph was quite parallel to the second ('a': 16%; 'b': 32%; 'c/d': 52%) (there was not a significant difference in the distributions between those two), different in admitting more predicted responses, but still mainly categorized with indeterminacy or complementary responses. However, that does give support to H90 uniquely, as it mainly provides further evidence that people reason with negative information differently than expected by the literature, and one possibility is that people consider sequential negations to be an affirmation, and in that case, a path counting model (which the literature does not advocate for path based reasoning) would also be predictive of those determinate responses.[2] To complete the set of comparisons, the first graph was compared with the third, and the difference in distributions was significant ($\chi^2(2) = 6.575, p < .05$). This, and the lack of significance in the distribution of responses between the second and third graphs, suggests that a 'redundant' link carries as much information as other links. This is consistent with the finding that 'preemptive' links do not have overriding status.

A comparison was also made with respect to a pair of graphs () and () symmetric to the first set, but the results differ somewhat with the original comparison and with the same comparison from Experiment One. Essentially, there was no significant difference in the distribution of responses to each of the problems (for the first — predicted, 'b': 13%; complement, 'a': 38%; indeter-

[2]Moreover, it's possible that a partition exists along these lines: those who see sequential negations as affirmations, those who see sequential negations as intensifiers, and those who see them as creating indeterminacy. A path counting model be parameterized for each of these perspectives.

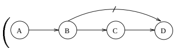
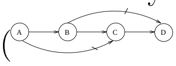
minate, 'c/d': 49%; and the second — predicted, 'b': 8%; complement, 'a': 39%; indeterminate, 'c/d': 53%). This is quite interesting since the literature predicts that the redundant link should make no difference. However, this is complicated by a few issues. Most important is that less than 15% of the people conformed to the literature prediction that the answer should have been 'b', favoring mainly indeterminacy and to a lesser extent, the complementary response. A second issue is that there was an effect of mode of presentation ($\chi^2(5) = 19.87, p < .01$). Under the graph only condition the results are essentially the same as for the sample as a whole, while under the sentence only condition there was a significant difference in the patterns of responses to the two graphs ($\chi^2(2) = 9.547, p < .01$). In the sentence only condition (as in the graph only condition) almost exactly half of the response was indeterminacy, but the remaining response was polarized: forty-six percent gave the predicted response 'b' to the first graph, and forty-three percent gave the complement answer 'a' to the second graph, contra to the literature prediction that answers should not switch between the graphs.[3] In sum, this comparison fails to support the literature prediction that redundant links should not complicate reasoning. The sentence only case gives further support for redundant links supplying as much information as any other. The significance of the mode of presentation effect is quite interesting since it suggests a sharper division between the way people reason with the same information presented graphically or sententially. The shift to the complement response was larger for this comparison than in the last one, and does replicate the trend observed in Experiment One.

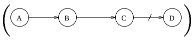
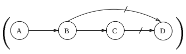
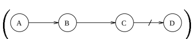
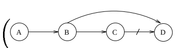
As for the first set of comparisons in this section, it is useful to compare the second graph (

) to a third related one that has a link of opposite polarity replacing the redundant one (

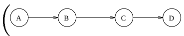
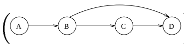
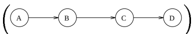
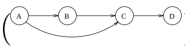
). In this case, the difference in distributions is significant (for the first — predicted, 'b': 8%; complement, 'a': 39%; indeterminate, 'c/d': 53%; and for the second — predicted, 'b': 30%; complement, 'a': 25%; indeterminate, 'c/d': 45%) over all ($\chi^2(2) = 13.924, p < .01$), as well as at both levels of the mode-of-presentation factor. The literature

[3] One potential explanation of this phenomenon is not that sentential presentation polarizes the response in itself, but that it makes the problems so much more difficult to reason about that when problems are sufficiently complex people who reach definite conclusions do so on the basis of the first chain of reasoning they are able to make through the sentences. Clearly in this connection the order of sentence presentation would have to be explored.

predicts ‘b’ as the response in both cases, and there is an increased proportion of ‘b’ responses for the third graph, though not significantly more than the number of complement responses. This result can be interpreted as additional evidence that people’s reasoning with negative information is quite distinct from the literature’s expectations: a path counting model that took into consideration the potential for negative chains to be used as negative paths could predict this trend in the data. The mode of presentation also had a significant effect on the association between the pattern of response and the presence of the redundant or nonredundant links. The difference comes mainly from the response to the third graph. In the graph only condition, the responses were mainly of indeterminacy (43%) with an exactly equal split of the remainder between the complement and predicted categories. In the sentence only condition, the response was more polarized (7% complement, 36% predicted, 57% indeterminate) towards the predicted answer. This finding is quite interesting given the ambiguous nature of the effect of reasoning with sentences and from graphs. The final relation to consider in this cycle is the comparison between the first () and third graphs (). In this case, the distributions of responses are significantly different ($\chi^2(5) = 23.25, p < .01$). The theoretical import of this is subsumed in the above discussion.

A final cycle of comparisons is made between the graph () and (), the latter of which contains a redundant link. Here, the literature defines the negative link between B and D in the second graph to be redundant; thus, the literature predicts a response of ‘b’ to both graphs. In fact, the observed response was overwhelmingly according to the literature prediction for both graphs: mainly the predicted answer (85%), some indeterminacy (13%), and a small amount in the complement category (2%). This contrasts with the related case reported in the preemption section: () and ().[4] Of course, there was also a significant difference in the patterns of response to the second and third graphs ($\chi^2(2) = 94.238, p < .01$). This cycle of comparisons

4 The difference in the distribution of responses to these two graphs is highly significant ($\chi^2(2) = 85.626, p < .01$). The first graph was categorized overwhelmingly as predicted (85%), somewhat as indeterminate (13%), and minimally with the complement category (3%), while the second was considered mainly indeterminate (68%), least with the predicted category (15%), and slightly more with the complement category (17%). From the first graph to the second, there is a tremendous shift towards indeterminacy and a substantial (15%) change in complement categorization, away from the predicted categorization.

substantiates the literature's predictions to a degree: since the findings suggest that where the addition of a topologically redundant link does not have any bearing on other conclusions, it is in fact redundant. This is quite different from the literature's notion of redundancy. An additional test of this hypothesis comes from comparing responses to the graphs () and (). Here, the literature predicts symmetrical results to those in the first pair of graphs in this cycle. And the trend towards predicted response is preserved, albeit with a just-significant ($\chi^2(2) = 7.774, p < .05$) shift towards the complement response (11% rather than 1%) in the graph with an additional link. The same trend also exists between the graphs () and ().

Summary. Responses to questions about networks that the literature deems to contain redundant links are shown in Table 3.6. The columns indicate the percentages of the participants who responded to that question by classifying it with the answer category indicated by the column heading (indeterminate = 'c/d', negative = 'b', positive = 'a'). The predicted response was negative classification for the first network, and positive for the second.

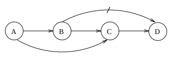
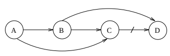
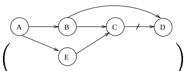
Redundancy				
Problem		Indeterminate	Negative	Positive
	Exp. 1	62	17	21
	Exp. 2	53	8	39
	Exp. 1	53	36	11
	Exp. 2	61	31	8

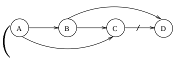
Table 3.6: Percentages of Responses in Each Answer Category: Are *As* typically *Ds*?

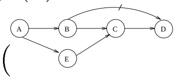
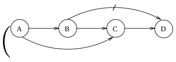
Off-path/On-path preemption

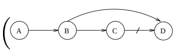
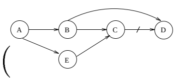
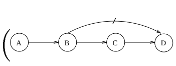
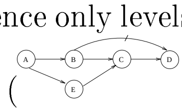
Recall that a distinction between varieties of preemption presuppose that preemption is an operative constraint. Both experiments found that graphs for which preemption would have made a difference did elicit significant differences in response patterns, but not in the direction predicted by the literature, suggesting that links topologically identified as preemptive actually supply only as much information as the other links. It is nonetheless interesting to compare graphs for which a distinction in types of preemption would make a difference, to simpler graphs in which preemption would also hold. A second style of comparison examines the potentially preempting graphs, contrasting them with other graphs that lack the explicit cancellation that preemption would supply.

Experiment One.

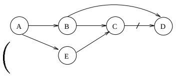
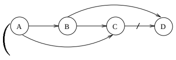
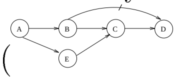
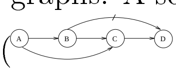
Two problems, (

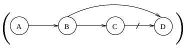
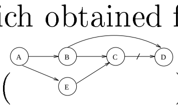
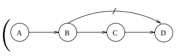
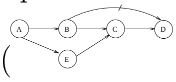
) and (

), were compared to test on-path vs. off-path preemption. The literature prediction for the latter problem was response 'a' (positive classification) under both definitions of preemption, and for the former 'a' according to off-path preemption but 'c/d' according to on-path preemption. Thus on-path preemption would predict a shift in responses to the two problems from 'a' for the latter towards 'c/d' in the former, whereas off-path preemption would predict no shift from response 'a'. In fact log linear analysis revealed that there was no difference in the distribution of subjects' responses to these two problems; however, this result cannot provide support for either theory of preemption since subjects responded mainly 'c/d' to both problems. The pattern of responses ran as follows: for the first problem – 'a' (off-path predicted) 9%, 'b' (complement) 32%, 'c/d' (indeterminate) 59%; for the second problem — 'a' 11% (predicted), 'b' 36% (complement), 'c/d' 53% (indeterminate). Even though H90 predicts no change in answer pattern between the two graphs, this result does not support that theory since the answer pattern for the base case (given in the second graph) confounds predictions to start with. The preponderance of 'c/d' answers in the second case is likely the result of some other factor like complexity of the problem leading to the conclusion that the graph is indeterminate, rather than something constrained by preemption simpliciter. There was a significant effect from mode of presentation on the association between the answer category and the presence or absence of the additional link. This effect was significant over all three levels of mode of presentation ($\chi^2(10) = 24.68, p < .01$) (where the trend was polarization with answers shared between the predicted and indeterminate categories in the graph only condition, wide distribution (but with focus on the complement and indeterminate categories) in the sentence only condition, and polarization shared between the complement category and indeterminate category in the graph+sentence condition) as well as over the graph only and sentence only modes alone ($\chi^2(5) = 19.82, p < .01$). A second, symmetric, comparison involved the graphs () and (). The literature prediction for the latter problem is response 'b' in both sorts of preemption. The predicted answer for the former problem differs: the off-path preemption prediction for that graph is 'b', whereas the on-path preemption prediction is 'c/d'. Again there was no significant difference in the distribution of subjects' responses to each of these problems (first problem: 'a', 17%; 'b', 27%; 'c/d', 56/regarding the relative plausibility of the two versions of preemption. There was no effect of mode of presentation in this case.

A second set of comparisons involved the graph () and () were compared. The literature prediction for the former is 'a' whether using on or off-path preemption. For the latter graph, the off-path preemption prediction is 'a', whereas the on-path preemption prediction is 'c/d'. For the first graph, the observed pattern of response was mainly indeterminacy ('c/d', 66%), and roughly equal amounts of 'a' (positive classification) and 'b' (negative classification) responses (15% and 19%, respectively). For the second, there was less indeterminate response (59% in total), more 'b' responses (32% in total) and fewer 'a' responses (9% in total). There was no significant difference in the distribution of subjects responses to each of these problems. However, since again the most frequent response to the former was 'c/d' and not 'a' as predicted, it is not possible for the results to distinguish between on-path and off-path preemption. There was not an overall significant effect of mode of presentation on the distribution of responses, but there was a significant effect at the graph only and sentence only levels ($\chi^2(5) = 15.72, p < .01$). Similarly, the graph () and () were compared. The literature predicts response 'b' for the latter graph using either on or off-path preemption, while for the former off-path preemption predicts 'b' whereas on-path preemption predicts response 'c/d'. There was no difference in the distribution of responses between these problems (the first graph: 'a', 22%; 'b', 24%; 'c/d', 54%; the second graph: 'a', 17%, 'b', 27%, 'c/d', 56%), nor was there a significant effect from the mode of presentation. Again since response 'c/d' was predominant in both cases it is not possible to distinguish between on-path and off-path preemption. Essentially, neither of these sets of comparisons is able to arbitrate between on-path and off-path preemption as a more plausible proof-theoretic device in modeling human reasoning. This seems to be the result of the complexity of the problems at stake — very likely other issues are determining the 'c/d' responses. Evidence that this is the case comes from the findings about the more foundational assumptions about redundant links and basic preemption built into inheritance reasoners.

Experiment Two.

The first comparison was between () and (). The literature prediction for the second problem was response 'a' (positive classification) under both definitions of preemption, and for the former 'a' with off-path preemption but 'c/d' (indeterminacy) with on-path preemption. There was not a significant difference in the distribution of responses to either of these graphs. There was a slight shift away from indeterminacy (61% reduced to 54%) towards answer 'a' (8% increased to 16%) between the first problem and the second, but the main responses were indeterminacy and the complement category. This was the basic finding in Experiment One as well. Note the substantial number of complement category responses to both problems (about 30% in each case): this would seem to be additional support for the predictions of a path-counting model since both graphs have two negative paths and a single positive path. The other trend is that the more complex the problem, the more indeterminate the response. There was an effect of mode of presentation approaching significance: essentially, in the sentence only condition virtually all of the responses were indeterminacy for both graphs. A second, symmetric, comparison involved the graphs () and (). The off-path preemption prediction for the first graph is 'b', whereas the on-path preemption prediction is 'c/d'. The literature prediction for the second graph is response 'b' in both sorts of preemption. There was only a very small difference in the pattern of responses to the two problems: about half of the participants rated both graphs indeterminate, only ten percent gave the predicted response, and the remaining forty percent answered with the complement response. The focus on the complement response occurred in the symmetric case as well, and strengthens support for path counting models since in this case there are two positive paths and only one negative path. Again no conclusions can be drawn regarding the relative plausibility of the two versions of preemption since preemption seems not to be a good predictor of responses in the first place. There was a just-significant effect ($\chi^2(5) = 14.14, p < .05$) of mode of presentation in that of the ten percent of the predicted responses to the first graph, sixty percent of them occurred in the sentences only condition.

A second set of comparisons examined whether there was a difference in response to the graphs for which on or off path preemption would make a difference and basic graphs without the extra chain of statements. These graphs were also used in testing basic preemption in the earlier section. Based on the lack of significance in the difference between response patterns to the complex and simpler preemptions in the preceding paragraph, one would expect the same results to hold in this set of comparisons which obtained for the basic preemption examinations. Consider () and (). The literature prediction for the former is 'a' whether using on or off-path preemption. For the latter graph, the off-path preemption prediction is 'a', whereas the on-path preemption prediction is 'c/d'. More than half of the response to both graphs was indeterminacy (68% for the former and 54% for the latter). For the first graph, the remainder of the response was nearly evenly divided (about 16%) between the predicted and indeterminate categories. About that same percentage responded with 'a' to the second graph as well, but about twice that many categorized the second graph using the complement response (the difference coming from the reduction in indeterminate responses). Again, this is supportive of a path-counting model as the second graph has two negative paths and only one positive path. However, this trend was not statistically significant. Neither was there an effect of mode of presentation. A final comparison in this section used graphs () and (), structurally similar to graphs used for a comparison in considering redundancy. As with the last comparison, and the one concerning redundancy, there was not a significant difference in the distributions of responses (for the first graph, 38%

'a' responses, 13% 'b' responses, and 49% 'c/d'; for the second, 41%, 12%, and 47%, respectively). The literature predicts 'b' for the first graph with either style of preemption, and also for the second graph in the case of off-path preemption. The on-path prediction for the second graph is 'c/d'. Again, the preponderance of indeterminate responses in both cases make it impossible to adjudicate between on-path and off path preemption. The lesser but nearly equal choice of the complement category for both graphs is rather interesting because apart from being contrary to the literature's prediction, it also is a surprise to the path-counting model, which would predict a difference between the two graphs. There was a significant effect of mode of presentation ($\chi^2(5) = 33.76, p < .01$) in which the trend actually went in the opposite direction under the sentence only condition (roughly equal proportions in the indeterminate and predicted categories and few in the complement). This gives more reason to believe that different strategies are in place for reasoning with sentential and graphical presentations.

Summary. Responses to questions about networks that differentiate reasoners with respect to on-path and off-path preemption are summarized in Table 3.7. The columns indicate the percentages of the participants who responded to that question by classifying it with the answer category indicated by the column heading (indeterminate = 'c/d', negative = 'b', positive = 'a'). The predicted response using off-path preemption was positive classification for the first network, and negative for the second. Using on-path preemption, the predicted response to the both problems is indeterminacy.

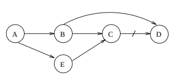
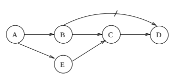
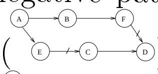
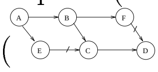
On-Path/Off-Path Preemption				
Problem		Indeterminate	Negative	Positive
	Exp. 1	59	32	9
	Exp. 2	54	30	16
	Exp. 1	56	27	17
	Exp. 2	47	12	41

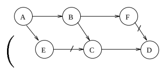
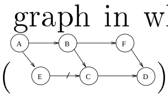
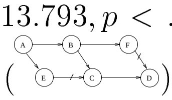
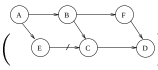
Table 3.7: Percentages of Responses in Each Answer Category: Are *As* typically *Ds*?

Cascaded ambiguities

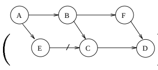
A final group of comparisons was intended to determine the most plausible approach to nested ambiguities.

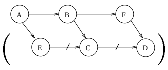
Experiment One.

The first minimal pair compares a graph that has just a negative path and a second chain that the literature would not define as a path () with another graph with a nested set of conflicting chains (). Inheritance reasoners tend to agree that people should say ‘b’ for the former problem, because there is only one proper path, and it is a negative path. H90 is distinct in its approach to the latter problem; if people behaved as predicted by H90 they should respond ‘b’ to the latter problem whereas if they followed the prediction of cascaded ambiguities in other reasoners they would respond with answer category ‘c/d’. Thus between the two problems H90 predicts no shift from response ‘b’, but other reasoners predict a shift from ‘b’ to ‘c/d’ responses. It was found that subjects did respond ‘b’ less (67% reduced to 31%) and ‘c/d’ more (26% increased to 58%) for the latter graph, as compared with the former (log linear analysis, $\chi^2(2) = 17.345, p < .01$) (the ‘a’ responses remained few — 7% and 11%). Thus this result challenges H90 and favors the prediction of other reasoners on this issue. There was not a significant interaction from the mode of presentation.

A second comparison examined one of the same graphs from above () with another graph in which there is a pair of conflicting paths embedded in a larger graph (). Both H90 and other reasoners predict response ‘a’ for the latter problem (as there are no conflicting paths between the extreme endpoints); however, H90 predicts response ‘b’ for the former problem whereas other reasoners predict response ‘c/d’ (as mentioned above). The observed response distributions ran thusly: for the first problem – ‘a’, 11%; ‘b’, 31%; ‘c/d’, 58%; for the second problem — ‘a’, 35%; ‘b’, 12%; ‘c/d’, 53%. The distribution of responses to these two problems were found to be significantly different (log linear analysis, $\chi^2(2) = 13.793, p < .01$); although subjects responded mainly ‘c/d’ to both, response to () was marked by a decrease in ‘a’ responses made to (

), accounted for by an increase mainly in 'b' responses but also in 'c/d' responses. This favors the H90 reasoner which predicts that a decrease in 'a' responses should be accounted for by an increase in 'b' responses (as opposed to 'c/d' responses as predicted by other reasoners), given that the increase in 'c/d' responses was not significant. However, the large number of 'c/d' responses (the dominant response) provides better support for non-H90 reasoners in general. Thus this result provides mixed evidence, and no clear support for either theory. There was not a significant interaction from the mode of presentation.

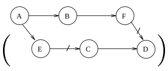
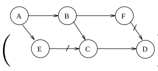
A final comparison looked at the second of the previous two problems (

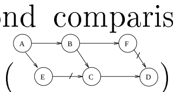
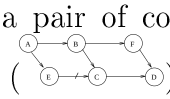
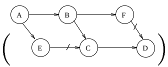
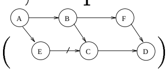
) which had an embedded pair of conflicting paths and compared it with a related one (

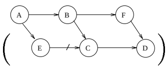
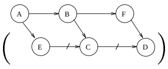
) which contains nested conflicting paths. There was a significant difference in the distribution of responses $\chi^2(2) = 8.95, p < .05$: 'a' responses decreased (from 35% to 16%) and 'b' responses increased (from 12% to 28%) between these two problems. The indeterminate response remained roughly constant for the two problems (53% and 56%, respectively). All reasoners predict that subjects should respond 'a' to the first problem, but whereas H90 predicts an 'a' response to the second other reasoners predict 'c/d' (meaning that the nested conflicts cascade into nested ambiguity). In fact, 'c/d' was the most frequent response (equally) to both problems (which, again, undermines support for H90 in general). There was a shift from 'a' responses as predicted by reasoners which cascade ambiguities, but this was accounted for by an increase in 'b' responses rather than the indeterminacy cascaded ambiguity would predict. The shift to 'b' responses is also contrary to the H90 prediction, but is compatible with a theory about negative chains being counted. There was not a significant interaction from the mode of presentation.

These results suggest a partitioning of respondents into the group that follows H90's prediction and the group that finds both graphs in the minimal pairs indeterminate. To the extent that such a partitioning is possible, there is some support for H90 through the second and third comparisons. However, this support is undermined by the first comparison which favored cascading ambiguities. The overwhelming percentage of 'c/d' answers perhaps indicates that when problems get complicated then they are increasingly likely to be found indeterminate, but the best metrics of complexity are difficult to identify. Alternatively, another explanation may be lurking in the data. Clearly more work needs to be done on this. It is interesting, however, to compare the significance of the results found in this section with the lack of significance in the preceding section: it does suggest that there is some systematicity to discover.

Experiment Two.

The first comparison was again between () and (), which differ in whether there is a nested ambiguity. The literature for the most part predicts the response 'b' for the former problem. H90 is distinct in its approach to the latter problem; if people behaved as predicted by H90 they would respond 'b' to the latter problem, but if following cascaded ambiguities they would respond with answer category 'c/d'. Thus between the two problems problems H90 predicts no shift from response 'b', but other reasoners predict a shift from 'b' to 'c/d' responses. As in Experiment One, it was found that subjects did respond 'b' less (by 51%) and 'c/d' more (by 36%) for the latter problem in comparison to their responses to the former, and the trend was highly significant (log linear analysis, $\chi^2(2) = 46.073, p < .01$). The association between the presence or absence of the nested ambiguity and answer category was quite strong (predicting the answer on the basis of the problem structure would have reduced the error of random guessing by 51% (using λ_y proportional reduction in error)). Thus, this result challenges H90 and favors the prediction of other reasoners on this issue. There was not a significant interaction from the mode of presentation.

As in Experiment One, the second comparison examined one of the same graphs from the preceding paragraph () juxtaposed with another graph in which a pair of conflicting paths is embedded in a larger graph that lacks a conflict (). Both H90 and other reasoners predict response 'a' for the latter problem (as there are no conflicting paths between the extreme endpoints); however, H90 predicts response 'b' for problem the former whereas other reasoners predict response 'c/d' (as mentioned above). The distribution of responses to these two problems were found to be significantly different (log linear analysis, $\chi^2(2) = 21.615, p < .01$); although subjects responded mainly 'c/d' to both, response to () was marked by a decrease in 'a' responses made to () (from 44% to 20%), accounted for by an increase mainly in 'b' responses (from 6% to 28%). Experiment One made a similar finding. There was not a significant interaction from the mode of presentation.

The final comparison was between the second of the preceding two networks () which had an embedded pair of conflicting paths (but no conflict at the endpoints) and () which contains nested conflicting paths, both internally and at the endpoints. There was a highly significant difference in the distribution of responses $\chi^2(2) = 8.95, p < .05$: 'a' responses decreased (from 44% to 6%), 'b' responses increased (from 6% to 18%) and 'c/d' responses increased (from 50% to 77%) between these two problems. The association between the presence or absence of the nested ambiguity and answer category was quite strong (predicting the answer on the basis of the problem structure would have reduced the error of guessing it at random by 35% (using λ_y proportional reduction in error)). Thus this result challenges H90. The literature agrees in predicting that subjects should respond 'a' to the first problem, but whereas H90 predicts an 'a' response to the second other reasoners predict 'c/d' (meaning that the nested conflicts cascade into nested ambiguity). In fact, 'c/d' was the most frequent response (as in Experiment One, though to a different degree) to both problems (which, again, undermines support for H90 in general). There was again an increase in 'b' responses as well, though the shift to 'b' was far less pronounced than the shift to 'c/d'. Neither H90 nor the extant systems which cascade ambiguities predict an increase in 'b' responses. These findings are also consistent with a path-counting model. In this case there was a significant effect of mode of presentation ($\chi^2(5) = 18.95, p < .01$) in that a smaller precentage rated the second graph indeterminate in the sentence only condition, preferring the answer category 'b'.

Summary. Table 3.8 summarizes responses to questions about networks that differentiate reasoners with respect to whether ambiguities in subpaths are cascaded. The columns indicate the percentages of the participants who responded to that question by classifying it with the answer category indicated by the column heading (indeterminate = 'c/d', negative = 'b', positive = 'a'). H90, which does not let ambiguity cascade, predicts that the first network should be classified negatively and the second one positively. Theories which disregard ambiguities in subpaths when considering conflicting superpaths would classify both networks as indeterminate.

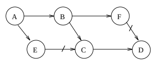
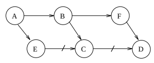
Cascaded Ambiguity				
Problem		Indeterminate	Negative	Positive
	Exp. 1	58	31	11
	Exp. 2	52	28	20
	Exp. 1	56	16	28
	Exp. 2	76	18	6

Table 3.8: Percentages of Responses in Each Answer Category: Are *As* typically *Ds*?

3.3 General Discussion

Experiment One finds human reasoning consistent with the inheritance literature in terms of its basic assumption of transitive reasoning with defaults, but found that people do not satisfy the literature's predictions about negative chains. More fundamentally, there was a tendency for people to take 'preemptive' links as additional information leading to indeterminacy rather than preemption as predicted by the literature. This renders adjudication between on-path and off-path preemption moot. On the other hand, there was support for the idea that links which express the same information that is contained in longer paths do not change responses to problems when added to graphs and are effectively 'redundant,' as the literature predicts. Under the assumption that it is possible to partition respondent categories, there was also support for H90's approach to nested ambiguities in a substantial minority. Graphic presentation of the problems polarized responses to problems in a way that suggests that the syntax of inheritance reasoners has influenced its proof-theory beyond its initial motivations for capturing human reasoning with generics. However, it is interesting to note that participants in the sentence only condition would occasionally make use of the space provided to work out their answers. Of the thirteen whose materials were returned with that space filled, or with accompanying paper, four subjects independently developed a graphic notation that was isomorphic to the inheritance graph that would have accompanied the corresponding problems; Elio and Pelletier (1993) found similar behavior. It could be argued that the mode of presentation effect which occurred here perhaps obtained because the sentence-only condition was assigned to a substantially younger population than the other twenty people tested. Experiment two addresses that criticism. Finally, the present analysis suggests that path-counting should also be studied in closer detail to determine its efficacy as a predictor of human reasoning with generics as a model of weighted evidence.

Table 3.9 summarizes the findings of both Experiment One and Two to facilitate an evaluation of inheritance proof theory. The numbers underneath the headings "Indeterminate," "Predicted" and "Complement" represent percentages of responses to each graph in those categories based on the total number of respondents who answered the question. Whether the predicted answer was the definite positive conclusion 'a' or the definite negative conclusion 'b' depends on the specific problem. The section of the table referring to the data on negative chains has slightly different headings since indeterminacy was the prediction in that case. "Negative" refers to answer category 'b' and "Positive," to 'a'. The predictions are those made by H90 as a theory of inheritance reasoning. The first three blocks of the table summarize the findings on transitivity, negative paths, and negative chains. The problem indicated is only a sample, as there were any number of longer or shorter problems than the one indicated; thus, the associated percentages are compiled over the appropriate set of problems in that case. The remainder of the results are the exact percentages accorded to the relevant problem depicted on the corresponding row. Results are indicated for both Experiment One and Experiment Two. Mode of presentation results are not included in this table because that aspect of the investigation is somewhat (although not entirely) orthogonal to the main issue of how people categorize the information implicit in a set of default sentences; instead, those results are briefly discussed later in this section. For each problem, the graph that represents it is given, rather than the set of sentences, because the graphs are the most visually concise index.

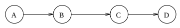
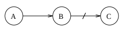
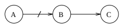
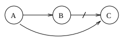
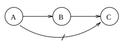
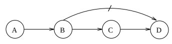
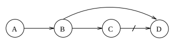
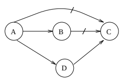
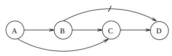
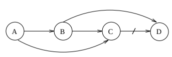
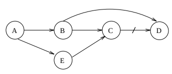
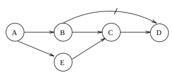
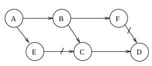
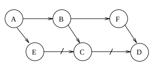
Transitivity				
Problem		Indeterminate	Predicted	Complement
	Exp. 1	23	66	11
	Exp. 2	8	87	5
Negative Paths				
Problem		Indeterminate	Predicted	Complement
	Exp. 1	27	65	8
	Exp. 2	11	86	3
Negative Chains				
Problem		Indeterminate (Pred)	Negative	Positive
	Exp. 1	47	41	12
	Exp. 2	19	75	6
Preemption				
Problem		Indeterminate	Predicted	Complement
	Exp. 1	42	43	15
	Exp. 2	54	38	8
	Exp. 1	43	49	8
	Exp. 2	62	25	13
	Exp. 1	54	24	22
	Exp. 2	49	13	38
	Exp. 1	66	15	19
	Exp. 2	68	17	15
	Exp. 1	37	53	10
	Exp. 2	52	43	5
Redundancy				
Problem		Indeterminate	Predicted	Complement
	Exp. 1	62	17	21
	Exp. 2	53	8	39
	Exp. 1	53	11	36
	Exp. 2	61	8	31
On-Path/Off-Path Preemption				
Problem		Indeterminate	Predicted	Complement
	Exp. 1	59	9	32
	Exp. 2	54	16	30
	Exp. 1	56	27	17
	Exp. 2	47	12	41
Cascaded Ambiguity				
Problem		Indeterminate	Predicted	Complement
	Exp. 1	58	31	11
	Exp. 2	52	28	20
	Exp. 1	56	28	16
	Exp. 2	76	6	18

Table 3.9: Percentages of Responses to Inheritance Networks

Table 3.9 makes some interesting facts clear. The most sharply defined results conforming to the literature predictions are in the top two chunks of the table. In both experiments subjects mainly behaved as predicted by the theories — that transitive reasoning is possible over positive paths, and that negative paths license a negative conclusion about the relationship between the endpoints of the path. The next result is also fairly sharply defined in Experiment Two: negative chains also license negative conclusions about the relationship between their endpoints. However, this result is contrary to the predictions of the literature which expected indeterminacy. Those results test rather foundational assumptions of the inheritance literature, rather than discriminating between particular proof theoretic proposals. Given the last of those findings a bit of doubt is cast upon the more complex aspects of proof theoretic debate — researchers have been arguing about the most intuitive ways to reason with highly complex graphs as if reasoning with theories that have the structure given by ($A \not\rightarrow B \rightarrow C$) is completely straightforward. Further exploration of this is required to ascertain how negative information is being used.

The next two sections of the table, on preemption and redundancy, are also foundational, but at a more conscious level of debate in the literature than the preceding issues. Assumptions about the nature of preemption and redundancy are what motivated Touretzky's (1986) original indictment of shortest path reasoning. Thus, these issues are closer to the ongoing debate in the proof theory of inheritance reasoning. However, they are still foundational since the literature assumes some topologies do involve preemption and redundancy; the argument is just about which ones. The results of Experiment One and to a larger extent, Experiment Two, indicate that mainly indeterminate responses were observed when the literature predicted some link to override a conflicting path and even more indeterminacy when the literature predicted a link to be ignored as redundant with respect to some path. In fact, this summarizes the results to the remainder of the table as well. For the more complicated problems the observed categorization was more than fifty percent 'indeterminate'. Interestingly, the remainder was split in such a way as to come closer to satisfying the predictions of the literature mainly with regard to simple preemption. Definite responses to problems involving redundancy and more complex forms of preemption tended to favor the complement response more than the predicted one. Definite response to problems involving nested ambiguities seems virtually random.

It is important to qualify these results with methodological considerations since, as mentioned above, the conditions were not balanced and in the first experiment subjects were not randomly assigned to the sentence only condition. The first point is not problematic given the sensitivity of log linear analysis. The second point does perhaps raise an issue since the Canadian pre-university students could for practical reasons participate in just the sentence only condition. Therefore, there was not a balance of subjects in the same age range in the the other conditions. For other uncontrollable reasons the second experiment was not balanced in terms of numbers of subjects in each condition, although it was balanced in terms of age and expertize of the students. This makes it interesting to consider the mode of presentation effects. Recall that in Experiment One the main effect was polarization of responses in the graph only condition, relative to the sentence conditions. Considering just the effect of the mode of presentation on category of answer there was a significant effect ($\chi^2(4) = 35.67, p < .01$) for all three levels. Basically, in the graph only condition there were roughly half as many definite responses ('a' or 'b' answers) as indeterminate ('c/d'), the definite answers shared equally between the two categories. In the sentence only condition there were one quarter more definite responses than indefinite responses. Twenty-one percent gave a definite positive classification, thirty-five percent gave a definite negative classification and forty-four percent gave an indeterminate classification. At the sentence+graph level there were again as many definite as indeterminate responses, however there were half as many definite positive classifications as there were definite negative classifications. Note that H90 would have predicted roughly equal numbers of answers in each of the three categories over the full questionnaire. Given that, and since at both levels of the mode of treatment that involved sentential presentations subjects were more likely to answer with the indeterminate or negative category, and further since the sentence+graph condition was balanced for ages, it seems reasonable to conclude that the age differences were insignificant. Rather, the conclusion might be that for whoever the audience the sentence condition is much more difficult to participate in than the graph only condition, and in that case a negative classification could be actually mistakenly applied as the denial of the positive classification (which is intended to be supplied by 'c/d'). Alternatively, some other reasoning strategy may be at work. Interestingly, the second experiment which was also balanced for age and background of subjects in the different modes of presentation (though unfortunately not in number) did not have an overall significant effect of mode of presentation on answer category (recall that these were undergraduates enrolled in composition or literature courses at a state university in Louisiana). In the sentence only condition of that experiment forty-three percent of the answers were in the indeterminate category, thirty-seven percent were negative, and twenty percent were positive, and virtually the same percentages existed for the graph only condition. Note that this is also nearly exactly the distribution that occurred in the sentence only condition of Experiment One. This gives additional reason to believe that the age unbalance in the first experiment was inconsequential (or it is an indictment of higher education in Louisiana).

Given these considerations, it is useful to reconsider the data with the goal of articulating a proof theory for inheritance that is in accord with human response to these problems. One thing that is clear is that classifications are sensitive to complexity. The more complex the problem the more likely the subject is to classify it as indeterminate, if that's an option. It would be worth investigating a condition in which there is more encouragement to reach definite conclusions, as in some sort of betting scenario or another situation that requires subsequent action on the basis of the conclusion. This does in fact correlate with a distinction made in the inheritance literature between skeptical and credulous reasoning. The results of the experiments reported here indicate that most people opt for an extreme skepticism when there is no need to act on the conclusions reached. In the context of having the option to be skeptical, the skeptics can be partitioned away to consider potential strategies for the remainder. One possibility that has been raised throughout the earlier discussion of the results is some version of path counting. Referring back to Table 3.9, note that the pattern of definite responses to problems testing redundancy and discriminations of preemption went in the opposite direction of the predictions of H90 as well as the rest of the literature (H90 predicted definite responses of one sort, and the rest of the literature expected indeterminacy). Recall from earlier discussion that each of those cases went in the direction that would have been predicted by a proof theory that advocates making a classification corresponding to the polarity of the greatest cardinality of paths through the network of sentences. Thus, some sort of weighting of amount of evidence seems most appropriate. The interesting thing from the point of view of the discussion within the inheritance literature is that this weighting would seem not to take into account particular kinds of sentences (when that kind is topologically defined, anyway) having preferred or inferior status — preemption by a shorter path is not allowed, although clearly some element in the partition of reasoners does behave in accordance with the predictions of preemption; topologically identified links are not disregarded as redundant. Another possibility, however, based on the distribution of negative and indeterminate classifications, has been suggested above — it is also possible that people use the negative classification as a form of denial of the positive one, even though the denial classification was an option, or that there is some other asymmetry in reasoning with positive and negative information. It would be useful for further work to be done (even using the data already amassed by Hewson and Vogel (1994) and this thesis) to study the combination of these strategies that produces a proof theory which has a 'best-fit' to the data.

This chapter has presented an experiment which provides data about human reasoning with generics and the degree to which human reasoning makes inheritance reasoning a plausible formal model. The focus has been on foundational assumptions of inheritance reasoning as well as some more controversial proof theoretic issues. The issues have been tested using abstract concepts and generic relations among them articulated in representative problems. Follow-up studies should investigate the interaction of less abstract interpretations. The present results factor out the difficult to control influence of personal knowledge and beliefs about real classifications like 'pacifist' or 'birds'. The inheritance literature would model the effect that specific background knowledge has on the conclusions derived from a set of generics involving concrete interpretations by encoding those beliefs as direct links and invoking preemption. Transitivity may accurately describe people's behavior in certain abstract cases as well as for interpreted instances in which inference is performed rather than direct recall: without further contextual information, people should reason transitively with the information represented in Fig. 3.5.a to conclude that penguins are fliers. The influence of world knowledge which might prevent this inference is modeled by inheritance reasoners with preemption by an explicit link, as represented in Fig. 3.5.b. However, the results presented here offer little support for preemption in ungrounded reasoning. Taplin and Staudenmayer (1973) cites the work of Wilkins (1928) as demonstrating definite effects of the degree of abstractness or concreteness on reasoning. Others have noticed effects of attitudes (Morgan & Morton, 1944; Kaufmann & Goldstein, 1967). Taplin and Staudenmayer (1973) also cites an argument of Henle (1962) that the existence of those effects does not invalidate formal logics as models of reasoning: she essentially makes the same claim that Hahn and Vogel (1995) point out about the possibility of formalizing legal knowledge and reasoning — the logic cannot be faulted for sanctioning conclusions that do not fit with context if the context has not been formalized in the first place.

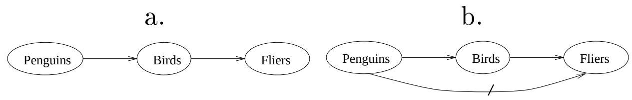


Figure 3.5: Representing Context

The nature of the conflicting intuitions on inheritance proof theory is such that it is informed more by studies that discriminate the influence of the abstract structure of a set of sentences than by the determination of which facts are explicitly represented and which are implicit. However, a more interesting set of contextual effects given the model of reasoning supplied by the inheritance literature would be those factors that have an impact on the abstract patterns of reasoning rather than the presence or absence of explicit information—for instance whether reasoning with known quantities (as opposed to using sets that have fuzzy cardinality, like 'chairs') eliminates the applicability of transitive reasoning. Studies of both abstract and grounded reasoning are important parts of the general problem of determining whether there is abstract systematicity in human reasoning with generic information.

Further studies should explore more of the proof-theoretic claims in greater detail. The initial result presented here on redundancy should also be examined further, especially given the conflicting intuitions in the literature about what the appropriate topological definition of redundancy should be (see Boutilier, 1989). For example, Horty et al. (1990) use a more complex definition than the one we suggested using informal terms in this paper; in theirs certain links that are redundant in the simpler terms are in fact deemed by them to convey novel information. We have not tested these problems. Finally, the present analysis suggests that path-counting should also be studied in closer detail to determine its efficacy as a predictor of human reasoning with generics as a model of weighted evidence.

Chapter 4

Parameterized Proof Theory for Inheritance Reasoning

4.1 Introduction

This chapter gives a declarative specification of a popular inheritance system and shows how simple changes to this specification can result in different path-based reasoners. This provides a deeper understanding of the fundamental differences between some of the more popular path-based inheritance reasoners. In particular, it allows the clarification of some of the results on the complexity of reasoning in the various systems. The declarative specification of H90 and its Prolog implementation was first presented by Vogel et al. (1993). That work set the foundations for parameterization of the proof theory into a uniform set of definitions which has not appeared before.

4.2 A Basic System

This section gives a logic based definition of the H90 system following Vogel et al. (1993). The plan is to define the notion of a permitted path, since it is the implicit link between the endpoints of a permitted path that expresses a conclusion of a

[0]Thanks to Fred Popowich of Simon Fraser University in British Columbia and Nick Cercone of the University of Regina for collaborations in this area, some of the material in this chapter, where indicated, is the fruit of that work.

reasoner with respect to a network.

4.2.1 Permission

The paths permitted in an arbitrary network by an inheritance reasoner are given here in terms of upwards construction of paths of increasing degree. It is also possible to state things in terms of downwards constructions (Touretzky et al., 1987) or double chaining (Touretzky, 1986). The formal definition follows:

Definition 8 (Permission)

1. Let π be a path.
   (a) If π is a direct link, then π is permitted.
   (b) If the degree of π is one then π is a direct link by the definition of degree, hence π is permitted.
2. Let π be a compound path of degree n. Assume that all permitted paths with degree less than n are known.
   (a) If π is a positive path then it has the form αz (i.e., $lastnode(\alpha) \longrightarrow z$ is a link in the network). The path α is positive, and by Proposition 1 in Chapter Two the degree of α is less than n. The path π is permitted iff
      i. α is permitted,
      ii. $firstnode(\alpha) \not\longrightarrow z$ is not a direct link in the net,
      iii. All negative paths $\pi' = \alpha'/z$, where α' is a permitted positive path (with degree less than n) and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.
   (b) If π is a negative path (it has the form, α/z), then π is permitted only under the conditions symmetric to those stated in (a). That is, iff:
      i. α is permitted,
      ii. $firstnode(\alpha) \longrightarrow z$ is not a direct link in the net,
      iii. All positive paths $\pi' = \alpha' z$, where α' is permitted positive path (with degree less than n) and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.

Definition 8 follows the inductive structure of the path-based definition for H90 despite the change from their network notation. Consistent with the H90 definition, the above definition proceeds with upwards construction of paths of increasing degree. According to these definitions, if conflicting paths intersect only at their endpoints, they will cancel each other. Cancellation is stipulated by condition (iii) on the permission of π in Definition 8. This condition states that π is permitted only if all conflicting paths are preempted. Since π has degree n, we know of all paths which could conflict with π. By the definition of degree, none of the conflicting paths is longer than n. The degree of a path becomes significant only during the examination of compound paths for the existence of conflicting paths. Conflicting paths are handled trivially in the case of direct links. All direct links are sanctioned as paths through a network, even conflicting links. If a direct link conflicts with a compound path then the definition of preemption is satisfied and the direct link preempts the compound path. In the case of conflicting compound paths, it is known that none of the conflicting paths has a degree greater than n, and all shorter paths between the same endpoints are known, because Definition 8 proceeds on the basis of increasing degree.

4.2.2 Preemption

Path preemption allows (more specific) information that is contained in a direct link to override conflicting information in a (more general) compound path. This topological ordering of paths is called the *inferential distance ordering* (Touretzky, 1986).

Only direct conflicting links can preempt other paths, although a preempting link may be part of a longer path.

Definition 9 (Preemption) Let π and π' be positive paths, and let p and y be nodes. A positive path πy is *preempted* by a link $p \not\longrightarrow y$ or a negative path π / y is *preempted* by a link $p \longrightarrow y$ if there exists a permitted path π' such that

1. $firstnode(\pi) = firstnode(\pi')$,
2. $lastnode(\pi) = lastnode(\pi')$,
3. p occurs in π', and
4. $p \neq lastnode(\pi')$.

When a path is preempted by a link $p \longrightarrow\!\!\!\!\!\circ\;\; y$, it is also said that the *node* p preempts the path. In the case where $p = lastnode(\pi')$ (i.e. where condition 4

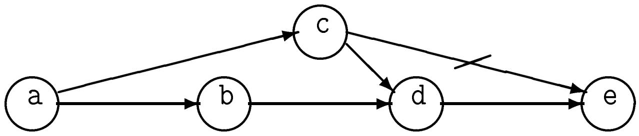


Figure 4.1: Path *abde* is off-path preempted by the link $c \not\rightarrow e$.

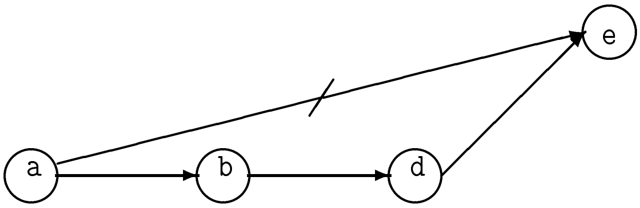


Figure 4.2: Link $a \not\rightarrow e$ preempts the path *abde*.

is violated), we would have two conflicting links $p \longrightarrow y$ and $p \not\rightarrow y$. Instead of allowing preemption, the network is said to contain an inconsistency.

An example of preemption is presented graphically in Figure 4.1, in which a positive path is preempted by a negative link. Matching the definition to the network in Figure 4.1, π corresponds to the path *abd*, π' to *acd*, p to c, and y to e. The endpoints of π and π' coincide, $p \not\rightarrow y$ is a link in the network, so *abde* is (off-path) preempted by the link $c \not\rightarrow e$. Essentially, ac/e is said to be more specific than both *acde* and *abde*. Another straightforward example of preemption is given in Figure 4.2: there is a similar matching between nodes and links of the network and the definition, except that in this case π and π' are identical.

Compound paths which conflict are subject to cancellation. The difference between cancellation and preemption is that neither conflicting path is permitted after cancellation, but preemption does permit one of its conflicting paths (namely, the path consisting of the direct link). Consider the network depicted in Figure 4.3 which has a topology similar to Figure 4.1. Although paths π and π' exist whose endpoints coincide, no π' exists that contains a node participating in a preempting link. Neither ac/e nor *abde* is favored over the other. In the network shown in Figure 4.1, preemption *resolves* an apparent ambiguity, but this does not happen in the network of Figure 4.3. Since this is a skeptical reasoner the paths cancel each other—neither is permitted. A credulous reasoner would resolve to two extensions from the network, one in which a's are e's and another in which

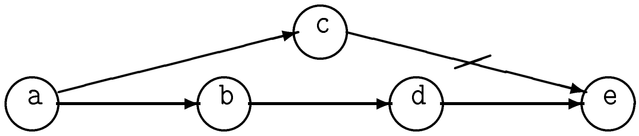


Figure 4.3: Paths ac/e and $abde$ cancel each other.

a's are not e's.

4.3 Other Systems

The formal definitions of path permission can be altered to provide alternative styles of inheritance reasoning. First it is helpful to restate Definition 8 in more schematic terms:

Definition 10 (Permission)

1. Let π be a path.
   (a) If π is a direct link, then π is permitted.
   (b) If the degree of π is one then π is a direct link by the definition of degree, hence π is permitted.
2. Let π be a compound path of degree n. Assume that all permitted paths with degree less than n are known.
   (a) If π is a positive path then it has the form αz. The path α is positive. The path π is permitted iff
      i. α is permitted,
      ii. $firstnode(\alpha) \not\longrightarrow z$ is not a direct link in the net,
      iii. no conflicting path matters.
   (b) If π is a negative path (it has the form, α/z), then π is permitted iff:
      i. α is permitted,
      ii. $firstnode(\alpha) \longrightarrow z$ is not a direct link in the net,
      iii. no conflicting path matters.

Definition 11 (Conflicts that Matter) Restricted Skepticism, Off-Path Preemption

No path conflicting with π matters iff

1. *if π has the form, αz, and all negative paths $\pi' = \alpha'/z$, where α' is a permitted positive path and $firstnode(\alpha) = firstnode(\alpha')$, are Off-Path preempted.*
2. *if π has the form, α/z, and all positive paths $\pi' = \alpha' z$, where α' is a permitted positive path and $firstnode(\alpha) = firstnode(\alpha')$, are Off-Path preempted.*

Off-Path preemption is exactly that sort which was specified in Definition 9. In the sections that follow I will modify this schematic definition to illustrate the relationships among the various proof theories for inheritance under consideration. The relationships among these styles of reasoning have not before been made explicit. Later in the chapter, a Prolog implementation is offered which has the advantage of perspicuous encoding since there is a direct textual relationship between the definitions given here and the Prolog code. The Prolog implementation of the Off-Path preempting, restricted skeptical reasoner (H90) was first presented by Vogel et al. (1993)

4.3.1 On-Path Preemption

Using on-path instead of off-path preemption, the nodes of the preempting link must occur on the path that is being preempted (as is the case depicted in Figure 4.2). On-path preemption is rather directly related to off-path preemption as just a more restricted form of preemption. Without changing the structure of Definition 9 we can define on-path preemption by adding the additional restriction that $\pi' \subseteq \pi$. Note that Vogel et al. (1993) erroneously state that to obtain on-path preemption, the relationship between π' and π should be equality. However, as discussed below, that restriction actually implements an alternative treatment of redundant links.

Definition 12 (On-Path Preemption) Let π and π' be positive paths, and let p and y be nodes. A positive path πy is *preempted* by a link $p \not\longrightarrow y$ or a negative path π/y is *preempted* by a link $p \longrightarrow y$ if there exists a permitted path π' such that

1. $\pi \subseteq \pi'$
2. $firstnode(\pi) = firstnode(\pi')$,
3. $lastnode(\pi) = lastnode(\pi')$,
4. p occurs in π', and
5. $p \neq lastnode(\pi')$.

Recall that reasoners which use off-path preemption conclude from the network in Figure 4.1 that As are not Es. On-path preemption leads to ambiguity in that network because c, the preempting node, does not occur on the path $abde$; thus that positive path conflicts with the negative path. Because preemption is about letting paths of opposite polarity override each other, using on-path preemption yields ambiguity where off-path preemption allows a definite conclusion.

4.3.2 Redundancy

Redundancy and Preemption

The literature assumes that certain topologically defined links are redundant and that a path through a set of redundant links should be preempted if the longer path with respect to which redundancy is determined is also preempted.

Definition 13 (Redundancy) *A path π is redundant with respect to a longer path π' if and only if they have the same polarity and:*

1. $\pi \subseteq \pi'$;
2. firstnode(π) = firstnode(π')
3. lastnode(π) = lastnode(π')

Intuitively, Definition 13 says that a path π is redundant with respect to π' if π is a subsequence of π' and they have the same endpoints. This just means that π contains direct links where π' has compound paths, so it is assumed that the direct links each contain no more information than is already in the corresponding longer path.

Thus, Definition 12 could have been given as follows:

Definition 14 (On-Path Preemption) Let π and π' be positive paths, and let p and y be nodes. A positive path πy is *preempted* by a link $p \not\rightarrow y$ or a negative path π/y is *preempted* by a link $p \longrightarrow y$ if there exists a permitted path π' such that

1. π is redundant with respect to π'
2. p occurs in π', and
3. $p \neq lastnode(\pi')$.

However, under the assumption that each link in a network conveys first-class information, preemption can still occur. Obviously, in a system without redundancy there is no need to stipulate preemption for paths that are redundant with respect to preempted paths. This means, in the above definition that π and π' are identical. Thus, on-path preemption for inheritance without redundancy reduces to the following:

Definition 15 (On-Path Preemption without Redundancy) Let π be a positive path, and let p and y be nodes. A positive path πy is *preempted* by a link $p \not\longrightarrow y$ or a negative path π/y is *preempted* by a link $p \longrightarrow y$ iff:

1. p occurs in π', and
2. $p \neq lastnode(\pi')$.

Using this definition, it is possible to conclude of the network in Figure 4.4 that *As* are not *Ds*, but not for the network in Figure 4.5. That is, the path *abcd* is still

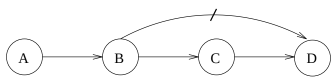


Figure 4.4: A Simple Inheritance Network

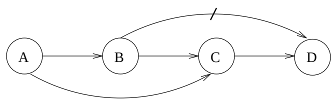


Figure 4.5: The Link $a \longrightarrow c$ Is Redundant

preempted, but the path *acd* is not since *b*, the preemptor, does not occur in *acd*.

Since *acd* and *ab/d* conflict, there is ambiguity, therefore it cannot be concluded that *As* are not *Ds*.

It is not coherent to define off-path preemption without redundancy since redundancy is actually just a special case of an off-path preemption, as the last example illustrates. This entails that in terms of relative numbers of paths deemed preempted, there is a total ordering such that off-path preemption is strictly more preemptive than on-path preemption which is in turn strictly more preemptive than on-path preemption without the assumption of redundancy. Eliminating preemption altogether is the only form of reasoning less preemptive than on-path preemption without redundancy. Thus, there are four possible forms of preemption that can be slotted into the definition of 'conflicts that matter'.

Redundancy and Permission

A final remark on redundancy is necessary in order to parameterize the definition of permission fully with respect to redundancy. Recall from Chapter Two that there are conflicting views (Boutilier, 1989) about how uniformly stable inheritance reasoners should be with respect to redundant links. For example, given the network in Figure 4.6, H90 would conclude *As are typically Es*, although if the network did not contain $c \longrightarrow e$ H90 would reach no conclusion about whether *As* are *Es* because in the smaller network the subpath *abcd* is not permitted since $a \overset{b}{\not\rightsquigarrow} d$, and its paths are assumed not to have nonfinal negative links. Nonetheless by Definition 13, $c \longrightarrow e$ is redundant with respect to the path *cde*. Boutilier (1989) argues that because $c \longrightarrow e$ is redundant, there should be no conclusion about whether *As* are *Es*, just as $a \longrightarrow c$ does not affect the conclusion that *As* are not *Ds*. He asserts that the presence of neither link should change the allowable conclusions. The other possibility, of course, is to accept the argument that both links be treated uniformly but to treat them both as non-redundant.

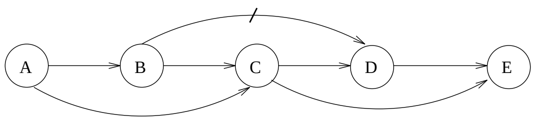


Figure 4.6: Are both $a \longrightarrow c$ and $c \longrightarrow e$ Redundant?

Definition 16 (Informational Redundancy) *A path π in a network ? is informationally redundant with respect to a longer path π' if and only if they have the same polarity and:*

1. $\pi \subseteq \pi'$;
2. firstnode(π) = firstnode(π');
3. lastnode(π) = lastnode(π');
4. $\forall \langle a, b \rangle$ *such that* $a \mathrel{-\!\!\ominus\!\!\!\rightarrow} b \in \pi, \exists \sigma$ *such that* $a \overset{\sigma}{\rightsquigarrow} b \in (? - (\pi \setminus \pi'))$

The definition of informational redundancy is more complicated than simple topological redundancy (Definition 13) in having an extra final condition which basically states that all of the links in the redundant path correspond to permitted paths (the implicit links) in the network even when reasoning is performed on the network minus those links. Using this definition and considering the network in Figure 4.6 again, it is clear that *ace* is redundant with respect to *abcde* (since both $a \overset{\emptyset}{\rightsquigarrow} c$ and $c \overset{\emptyset}{\rightsquigarrow} e$ are implicit links, permitted even when the network doesn't contain them explicitly). Using H90 on the network shown in Figure 4.7, the path fae is informationally redundant with respect to both $face$ and $fabcde$. However, using Boutilier's (1989) system fae is informationally redundant with respect to neither $face$ nor $fabcde$; this is because his system does not sanction a permitted path between a and e. Therefore in the network without the explicit link $a \longrightarrow e$ there will not be an implicit link between the two nodes either. Clearly, informational redundancy ignores fewer paths than redundancy simpliciter.

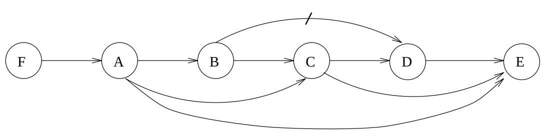


Figure 4.7: Is $a \longrightarrow e$ Redundant?

It is possible to amend the definition of conflicts that matter when chaining a path together to determine if it is permitted. Definition 17 incorporates the alternative view of redundancy of Boutilier (1989), but casts his system into the terms used within this thesis rather than utilizing the double-chaining mechanism that he used to achieve the same effect.

Definition 17 (Conflicts that Matter: Boutilier (1989)) *No path conflicting with π matters iff*

1. *if π has the form, αz and there is no unpermitted path with respect to which αz is informationally redundant, then all negative paths $\pi' = \alpha'/z$, where α' is a permitted positive path and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.*
2. *if π has the form, α/z and there is no unpermitted path with respect to which α/z is informationally redundant, then all positive paths $\pi' = \alpha' z$, where α' is a permitted positive path and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.*

The apparent circularity between the definitions of informational redundancy and the conflicts clause of the preemption definition is not vicious because the appeal to preemption in the informational redundancy definition is with respect to a smaller graph which has links under scrutiny removed. Informational redundancy is still a topological definition — it employs only a very weak form of information and is sensitive to the other constraints on reasoning within some system. That is, it basically asserts that a link which is redundant simpliciter may not be redundant when considered by a reasoner which would not have actually licensed an implicit link between the endpoints of the longer path. If the longer path is not permitted, then a link which spans its endpoints actually does add information to the network. However, this is still quite a weak form of information to consider. Given the network in Figure 4.6, both $a \longrightarrow c$ and $c \longrightarrow e$ are informationally redundant, but since H90 differentiates them, ignoring $a \longrightarrow c$ to conclude $a \overset{b}{\not\rightsquigarrow} d$ but regarding $c \longrightarrow e$ as evidence for the conclusion $a \overset{bc}{\rightsquigarrow} e$, there is reason to feel there should be an intuitive way of defining redundancy such that $a \longrightarrow c$ is redundant but $c \longrightarrow e$ is not. As it stands, both links are informationally redundant, and certain topological configurations that the redundant link can exist in have an impact on whether the reasoner is stable.[1]

It should be clear that there is no point in integrating informational redundancy into off-path preemption, because the notions are antithetical (the basic idea of off-path preemption being that even compound paths, essentially 'redundant links' with intervening nodes can be redundant). Integrating informational

[1]Recall that stability is the property of a reasoner reaching the same conclusions with or without redundant links explicitly present.

redundancy into on-path preemption, as in Definition 18 yields a quite skeptical system. Essentially, there will not even be preemption as in Figure 4.1 since π and π' are identical in that case, and a lone path cannot be informationally redundant (if its links were removed there would be no other path to license them as implicit links). While paths like *acd* in Figure 4.5 are preempted using this definition, by virtue of the fact that *acd* is informationally redundant with respect to *abcd*, the path *abcd* is not preempted (and therefore the network is ambiguous). The difference between this and the previous application of informational redundancy is that here it is integrated into the definition of preemption, but there it was just used as a check on allowable conclusions. Perhaps it seems futile to define a version of on-path preemption which does not actually admit preemptions in the usual sense. However, it is useful to have this system as a point in the spectrum between considering all paths as containing first class information (as in some scheme in which numbers of paths of each polarity determine the final conclusion) and other systems in which preemptions occur without regard to the number of conflicts. That is, the systems defined so far have used preemption to reach definite conclusions and assumed that some paths can be discounted, disregarding the numbers of each; the system just defined still discounts some paths as well as numbers in conflicting polarities, but it does not preempt anything to reach a definite conclusion; systems defined later will not discount any path, will consider relative cardinalities and will reach definite conclusions where possible.

Definition 18 (Preemption with Informational Redundancy) Let π and π' be positive paths, and let p and y be nodes. A positive path πy is *preempted* by a link $p \not\longrightarrow y$ or a negative path π / y is *preempted* by a link $p \longrightarrow y$ if there exists a permitted path π' such that

1. π is informationally redundant with respect to π'
2. p occurs in π', and
3. $p \neq lastnode(\pi')$.

Summary

This section has outlined three forms of redundancy which exist in a decreasing order of dismissiveness of paths. Redundancy in H90 effectively includes certain compound paths through the way off-path preemption is specified. On-path preemption utilizes a slightly tighter form that labels paths as redundant solely if they contain redundant links, where redundant links are just those that span a longer path of the same polarity. Informational redundancy is a more restricted version in which the paths spanned by redundant links must be permitted. Incorporating informational redundancy into preemption yields a form of preemption which preempts only redundant paths but not the longer paths with respect to which the smaller ones are redundant. Because preemption is invoked only in contexts where there are paths of conflicting polarity, this implies that using preemption with informational redundancy will yield fewer definite conclusions than preemption based on strictly topologically defined preemption. Preemption without the assumption of redundancy leads to a similar state of affairs. Figure 4.8 illustrates the consequences of fitting these various definitions together in terms of the paths preempted by the various systems. H90 preempts the most paths, and hence more often reaches definite conclusions (this of course is dependent upon the network reasoned about, for some networks all four will reach the same conclusions). Using on-path preemption instead of off-path preemption in H90 yields a system in which only direct links (and the paths comprised of them) can be redundant—such a system preempts strictly more paths than either on-path preemption without the assumption of redundant links and on-path preemption with informational redundancy. The last two systems are not in a subsumption relation as has been seen in the examples above: the former still allows paths to be preempted when there are no 'redundant links' but yields ambiguity in other cases, and the latter preempts paths comprised of redundant links, but does not allow paths to be preempted that do not contain redundant links, yielding ambiguity.

A path π in a network ? is redundant with respect to a longer path π' if and only if they have the same polarity and:	
Path/Link Redundancy:	$\mathit{firstnode}(\pi) = \mathit{firstnode}(\pi')$, and $\mathit{lastnode}(\pi) = \mathit{lastnode}(\pi')$;
Link Redundancy:	$\pi \subseteq \pi'$, and $\mathit{firstnode}(\pi) = \mathit{firstnode}(\pi')$, and $\mathit{lastnode}(\pi) = \mathit{lastnode}(\pi')$;
Informational Redundancy:	$\pi \subseteq \pi'$, and $\mathit{firstnode}(\pi) = \mathit{firstnode}(\pi')$, and $\mathit{lastnode}(\pi) = \mathit{lastnode}(\pi')$, and $\forall \langle a, b \rangle$ such that $a \multimap\!\!\rightarrow b \in \pi, \exists \sigma$ such that $a \overset{\sigma}{\rightsquigarrow} b \in (? - (\pi \setminus \pi'))$.

Recapitulating Redundancy.

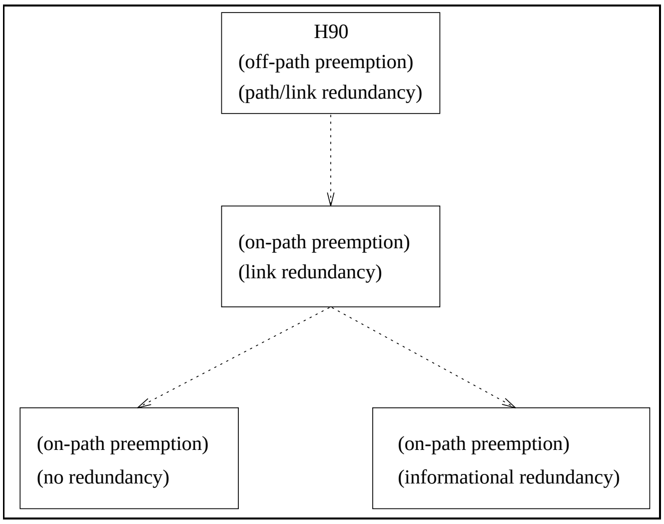


Figure 4.8: Paths Preempted by Various Systems

4.3.3 Skepticism

Fully Skeptical Reasoning

Recall the network which differentiates restricted skeptical and fully skeptical reasoners. For convenience, this figure is reproduced in this section as Figure 4.9.

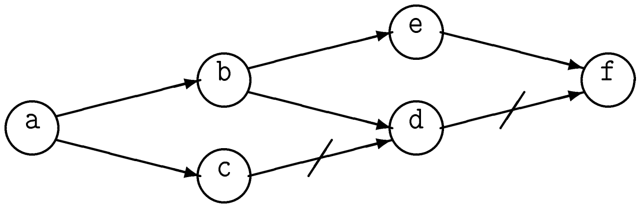


Figure 4.9: A Network with Potentially Cascaded Ambiguity.

The reasoner defined and implemented in the previous section was a restricted skeptical reasoner: ambiguities do not cascade through a network. Restricted skepticism is independent of a choice of preemption methods. In the network of Figure 4.9 we can conclude that a is f because there is no path permitted which conflicts with $abef$. The chain of links abd/f, which appears as if it could conflict, is not actually permitted because abd is not permitted, and it is impossible to build a path by extending a chain that is not itself a permitted path. In Definition 8, this was stipulated inside condition (iii) which determines what it means for a conflicting path to matter (Definition 11): α', part of the conflicting path π' must be a variable over positive paths.

A fully skeptical reasoner demands that ambiguities cascade through a network. Thus, a fully skeptical reasoner should reach no conclusion about a's f-ness, because a non-preempted negative path conflicts with the positive path $abef$. In the definition, α' is a variable over positive paths without regard for whether α' is actually permitted, so long as it is not preempted by more specific information.

Definition 19 (Conflicts that Matter: Subpath Credulity) *No path conflicting with π matters iff*

1. *if π has the form, αz, and all negative paths $\pi' = \alpha'/z$, where α' is a positive path and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.*
2. *if π has the form, α/z, and all positive paths $\pi' = \alpha' z$, where α' is a positive path and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.*

This difference is orthogonal to the question of which form of preemption gets used, therefore the specification of which conflicts matter is referred to as *subpath credulity*. That is, the restricted skeptical reasoner is skeptical about ambiguous subpaths, and ignores them. A reasoner that uses subpath credulity will assume that ambiguous subpaths are actually positive when considering the potential ambiguity of longer paths. The effect is that ambiguity cascades through the network to the longer path, and in the end, it's the subpath credulous system that concludes indeterminacy more than the restricted skeptical reasoner. For want of a better term this is called fully skeptical reasoning, but since either form of preemption can be used with subpath credulity and since on-path preemption will result in more indeterminacy than off-path preemption, there is strong motivation for calling a fully skeptical reasoner that uses on-path preemption even more skeptical than one that uses off-path preemption.

Ideally Skeptical Reasoning

One good reason for calling fully skeptical reasoning, reasoning in which ambiguities cascade, something other than "fully skeptical" is that it is not coextensive with ideally skeptical reasoning, even though it seems to be at first glance. Ideally skeptical reasoning is taken to be the intersection of credulous expansions (see Stein, 1989). However, there can be elements of the intersection of all credulous expansions of a network that are licensed by unique paths which do not occur in all credulous expansions. For example, Stein (1989) provides a network as in Figure 4.10 with the following interpretation to the nodes: $a =$*seedless grape vine*, $b =$*grape vine*, $c =$*infertile thing*, $d =$*fruit plant*, $e =$*vine*, $f =$*arbor plant*, $g =$*tree*, $h =$*plant*. Stein (1989, p.1156) points out, "Whether a seedless grape vine is a fruit plant or an arbor plant, it is certainly a plant!" That is, $a \rightsquigarrow h$ is an element of each credulous expansion. However, it is by virtue of $a \overset{bef}{\rightsquigarrow} h$ in the expansions in which seedless grape vines are not fruit plants (ac/d is the permitted path, hence abd/f isn't available to conflict with the subpath $abef$), and by virtue of $a \overset{bdg}{\rightsquigarrow} h$ in the expansions in which seedless grape vines are fruit plants. Restricted skepticism does get the desired conclusion here, but does not correspond to ideally skeptical reasoning as discussed in Chapter Two (because it licenses some conclusions that ideally skeptical reasoning would not). Surprisingly at first glance, fully skeptical reasoning does not capture ideally skeptical reasoning either, as can be seen from the fully skeptical expansion of the network shown in Figure 4.10: because the subpath abd is not permitted (due to ambiguity), the path $abdgh$ is not permitted; because the subpath $abef$ is not permitted (due to ambiguity), the path $abefh$ is not permitted. Thus, there is not a way using the fully skeptical reasoner to built a path that licenses the implicit link $a \rightsquigarrow h$.

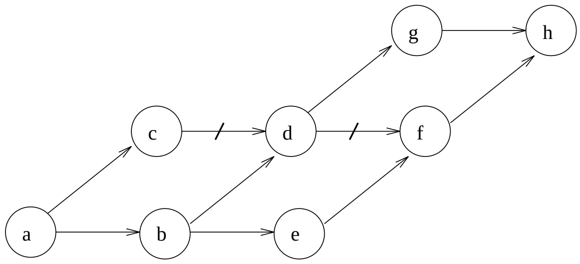


Figure 4.10: Ideally Skeptical Reasoning: $a \rightsquigarrow h$

Stein (1989, p.1156) concludes:

> There are facts which are true in all credulous extensions, but which have no justification in the intersection of those extensions. This is why we cannot generate a "skeptical extension"—no *particular* set of edges of ? from *seedless grape vine* to *plant* is in every credulous extension, so no such path can be in the "skeptical extension." Thus every path-based approach to skeptical inheritance will always be either unsound or incomplete with respect to ideally skeptical inheritance. We can only compute the always-true inferences by, in effect, reasoning about all of the credulous extensions. Fortunately, in acyclic hierarchies, such reasoning is tractable.

Stein (1989) proceeds to build a truth-maintenance system which does reason about all of the credulous extensions.

Note that in spite of the subpath ambiguity it is possible even in the fully skeptical reasoner to conclude $e \overset{f}{\rightsquigarrow} h$ and $d \overset{g}{\rightsquigarrow} h$. That is, the property of being a plant inherits all the way down to node b, regardless of the ambiguity between b and f. The only reason it doesn't inherit all the way to a, of course, is because of the subpath ambiguity between a and d. However, note that from the vantage point of node b, if h is inherited all the way to d and to e, then even if there is a subpath ambiguity about whether $b \rightsquigarrow f$ holds it is nonetheless reasonable to conclude that $b \rightsquigarrow h$, since b can get to h through either d or e. Contrast this with the network in Figure 4.11, which is just a subset of the one from Figure 4.10. The smaller network does license $e \overset{f}{\leadsto} h$, however there is no other information about h to reason with in the credulous expansion that contains $b \overset{d}{\not\leadsto} f$. Essentially, all that is required is an inheritance version of reasoning by cases to justify $b \leadsto h$ in the first example, and that principle's inapplicability in the second case distinguishes the two. In the absence of an explicit preemption between a and h and given that $b \leadsto h$ and $a \overset{\emptyset}{\leadsto} b$, it is reasonable to conclude $a \leadsto h$. Moreover, it is reasonable to even associate that conclusion more directly with a path: $a \overset{bef}{\leadsto} h$. This is because while $abdgh$ suffers from subpath ambiguity, the fact that $e \overset{f}{\leadsto} h$ still gets information from h to b, therefore $abefh$ is the empowering path. Admittedly, this is a weak relationship, since it too suffers subpath ambiguity, but that ambiguity is already overridden by the case-reasoning.

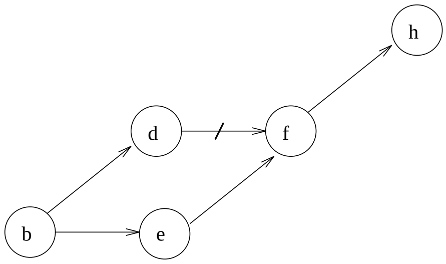


Figure 4.11: Ideally Skeptical Reasoning: $\neg\exists(b \leadsto h)$

Definition 20 (Conflicts that Matter: Subpath Credulity + Cases) *No path conflicting with π matters iff*

1. *if π has the form, αz, and all negative paths $\pi' = \alpha'/z$, where α' is a positive path and $firstnode(\alpha) = firstnode(\alpha')$ are (i.) preempted or (ii.) for each node i in α where $i \neq$ firstnode(α), $\exists\omega$, $i \overset{\omega}{\leadsto} z$.*
2. *if π has the form, α/z, and all positive paths $\pi' = \alpha' z$, where α' is a positive path and $firstnode(\alpha) = firstnode(\alpha')$, are (i.) preempted or (ii.) for each node i in α where $i \neq$ firstnode(α), $\exists\omega$, $i \overset{\omega}{\leadsto} z$.*

Definition 20 specifies this new form of skeptical reasoning, which implements reasoning with cases, here in the context of cascaded ambiguity. Essentially, the only difference is that for all paths that conflict with the subpath under extension (into a longer chain), each node but the first one already permits the endpoint of the chain. Thus, this definition incorporates reasoning by cases into skeptical path-based reasoning, and I argue that it is complete with respect to ideally skeptical reasoning. Stein's point, quoted above, still stands, since by sanctioning paths that yield the implicit link as permitted, it licenses paths that are not contained in all credulous expansions, and is thus not sound with respect to ideally skeptical reasoning. It would be possible to generalize the system along the lines of Veltman's (as discussed in Chapter Two), to distinguish syntactically between explicit and inferred links, for instance by incorporating abstraction over paths in between endpoints ($\rightsquigarrow$). Thus far I have been using the symbolism $i \rightsquigarrow j$, without explicit mention of a path as in $i \overset{\pi}{\rightsquigarrow} j$, as an abbreviation for "a path permitted between i and j". Definition 21 formalizes this, but notice that it does not have an 'only-if' condition: a definition of *weak permission* (see Definition 22) supplies the other conditions on the weak permission of an implicit link between the endpoints of a path that is not itself permitted. If this sort of distinction is made use of in defining inheritance, then it is straightforward to license the conclusion corresponding to the endpoints of the path without actually requiring support of the path itself, as in Definition 22. This achieves soundness with respect to ideally skeptical reasoning as well, although it is in a system slightly more expressive than path-based reasoning.

Definition 21 (Implicit Link Abstraction) *Given a network ? and a reasoner $\mathcal{R}$:*

- $x \rightsquigarrow y$ *if* $\exists \pi,\ x \overset{\pi}{\rightsquigarrow} y$
- $x \not\rightsquigarrow y$ *if* $\exists \pi,\ x \overset{\pi}{\not\rightsquigarrow} y$

Definition 22 (Ideally Skeptical Permission ($\rightsquigarrow$))

1. Let π be a path.
   (a) If π is a direct link, then π is permitted ($x \overset{\emptyset}{\rightsquigarrow} y$, where *firstnode*$(\pi) = x$ and *lastnode*$(\pi) = y$).
   (b) If the degree of π is one then π is a direct link by the definition of degree, hence π is permitted ($x \overset{\emptyset}{\rightsquigarrow} y$, where *firstnode*$(\pi) = x$ and *lastnode*$(\pi) = y$).

2. Let π be a compound path of degree n, with *firstnode*$(\pi) = x$. Assume that all permitted paths with degree less than n are known.
   (a) If π is a positive path then it has the form αz. The path α is positive. The path π is permitted ($x \overset{\alpha}{\rightsquigarrow} z$) iff
      i. α is permitted,
      ii. $firstnode(\alpha) \not\longrightarrow z$ is not a direct link in the net,
      iii. no conflicting path matters (in the fully skeptical sense).
   (b) If π is a negative path (it has the form, α/z), then π is permitted ($x \overset{\alpha}{\not\rightsquigarrow} z$) iff:
      i. α is permitted,
      ii. $firstnode(\alpha) \longrightarrow z$ is not a direct link in the net,
      iii. no conflicting path matters (in the fully skeptical sense).
   (c) If π is a positive path then it has the form αz. The path α is positive. The path π is weakly permitted ($x \rightsquigarrow z$) iff
      i. α is at least weakly permitted ($x \rightsquigarrow$ *lastnode*(α)),
      ii. $firstnode(\alpha) \not\longrightarrow z$ is not a direct link in the net,
      iii. no conflicting path matters (in the extended skeptical sense).
   (d) If π is a negative path (it has the form, α/z), then π is weakly permitted ($x \overset{\alpha}{\not\rightsquigarrow} z$) iff:
      i. α is at least weakly permitted ($x \rightsquigarrow$ *lastnode*(α)),
      ii. $firstnode(\alpha) \longrightarrow z$ is not a direct link in the net,
      iii. no conflicting path matters (in the extended skeptical sense).

While Definition 22 does capture ideal reasoning, it does so by utilizing abstractions over paths (using the network in Figure 4.10 as an example, it does not license $a \overset{bef}{\rightsquigarrow} h$, but it does license $a \rightsquigarrow h$), and is therefore not strictly a path-based inheritance reasoner, yet it does this without reasoning about all credulous expansions. However, I have argued that there is reason to feel the paths licensed by the extended skepticism (with both subpath credulity and reasoning with cases) are in fact supportable, even though they do not occur in all credulous expansions (note, for example that it only licenses some of the paths whose endpoints are represented in the implicit link). For the remainder of this thesis I will focus on strictly path based accounts. I will not provide an implementation of ideally skeptical reasoning, but will implement full skepticism extended with reasoning by cases.

A reasoner that uses subpath credulity and reasoning by cases will not cascade ambiguity to the same extent that the unextended skepticism will, however it will in many cases reduce to exactly the same thing. Just as in skeptical reasoning with subpath credulity, adding reasoning about cases is compatible with both forms of preemption. However, because it licenses conclusions about paths (which in ideal skepticism we can call weakly permitted) that have ambiguous subpaths, the extended skepticism yields less ambiguity than fully skeptical reasoning.

Summary

This section has presented a series of skeptical reasoners with different degrees of skepticism. The literature makes its sharpest distinction between credulous reasoning in which ambiguities are resolved one way or another in two or more expansions and skeptical reasoning in which there is always just one expansion (and faced with a genuine ambiguity, no determinate conclusion is made). Clearly, though, there is more than one kind of skepticism to work with. To fit with traditional notions of skepticism, it makes sense to consider the possibilities in terms of their relative capacity for classifying a network as ambiguous. H90 reaches a definite conclusion when at all possible. Allowing ambiguities to cascade (requiring subpath credulity) gives the opposite end of the spectrum in which a network is more likely to be considered ambiguous. The two possibilities in between are subpath credulity augmented by reasoning with cases or ideally skeptical reasoning. The preceding section described how ideally skeptical reasoning licenses conclusions that correspond to no path that exists in the intersection of all credulous expansions. The other system introduced here, cascaded ambiguities with case reasoning, approximates ideally skeptical reasoning, but is not identical since it licenses paths as well as the implicit links they justify. A summary of these relationships is depicted in Figure 4.12. The variation discussed in the next section, downwards reasoning, is orthogonal to skepticism associated with paths and subpaths, and for each combination of the preceding systems yields a system that is more likely to find ambiguity.

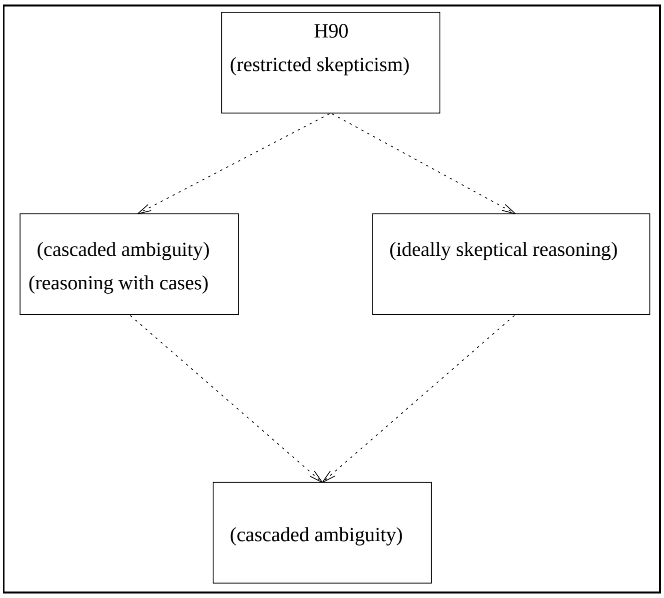


Figure 4.12: Paths Preempted by Various Systems

4.3.4 Direction of Reasoning

Recall from Chapter Two that the direction of chaining also affects the conclusions that are sanctioned by a reasoner. Chapter Two surveyed a few approaches to direction of chaining, and pointed out that the negative complexity results are attached to a form of double chaining. H90, a polynomial system, utilizes forward single chaining, but both Touretzky (1986) and Boutilier (1989) rely on double chaining. The complexity results of Selman and Levesque (1989, 1993) for downwards reasoning actually apply to double chaining. Actually, downward reasoning need be no more complex than upwards chaining, which fits with the observation also pointed out in Chapter Two that a downwards reasoner will never resolve to a single expansion where an upwards reasoner would find ambiguity (ambiguities are easier to identify than definite conclusions).

Definition 23 (Downwards Permission)

1. Let π be a path.
   (a) If π is a direct link, then π is permitted.
   (b) If the degree of π is one then π is a direct link by the definition of degree, hence π is permitted.

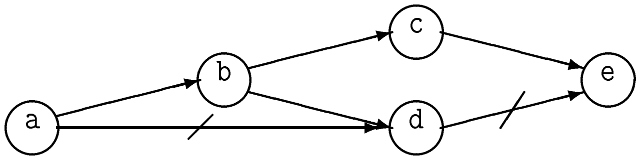


Figure 4.13: Preemption by $a \not\rightarrow d$ Permits the Conclusion $a \overset{bc}{\leadsto} e$.

2. Let π be a compound path of degree n. Assume that all permitted paths with degree less than n are known.
   (a) If π is a path then it has the form $x \longrightarrow \alpha$. The path α may be positive or negative. The path π is permitted iff
      i. α is permitted,
      ii. $x \mathrel{-\!\!\ominus\!\!\rightarrow} \mathit{lastnode}(\alpha)$ is not a direct conflicting link in the net,
      iii. no conflicting path matters.

Note that this definition assumes the standard definition of path construction: a negative link can occur only as the last link of a path. This makes it easier to collapse the specification of permission of positive and negative paths. Just as in the forward chaining definition of permission, the path being chained onto has to be allowable, it's just that in this case it's the back half of the path rather than the front half. Everything else remains the same. This is sufficient to obtain the difference in reasoning about the network depicted in Figure 4.13: even though the path abd/e is preempted in an upwards reasoner, thus allowing the conclusion $a \overset{bc}{\leadsto} e$, for a downwards reasoner it is not possible to chain past the ambiguity at b about whether $b \overset{c}{\leadsto} e$ or $b \overset{d}{\not\leadsto} e$, and therefore no conclusion is reached about whether there is an implicit link between a and e. The various forms of preemption can also be slotted into downwards reasoning.

4.3.5 What People Really Use

At this point, definitions have been provided for twenty-one reasoners,[2] although they are not all distinct. The possible combinations are indicated in Table 4.1. There is overlap among the twenty-one: for instance, off-path preemption with

[2]Actually, forty-two if you include upwards and downwards chaining.

only link redundancy is just on-path preemption. The advantage of setting the reasoners out in these terms is that it lays the foundation for inspecting the relationships among them so that essential equivalences and differences can be identified.

Preemption		Degree of Skepticism		
		Restricted	Fully	Extended
On-Path				
Redundancy		✓	✓	✓
path/link	×			
link	✓			
informational	✓			
none	✓			
Off-Path				
Redundancy		✓	✓	✓
path/link	✓			
link	✓			
informational	✓			
none	✓			

Table 4.1: A Space of Possible Default Inheritance Reasoners

Recall that Hewson and Vogel (1994) and Chapter Three tested people for patterns of response to sets of sentences and compared the observed responses with the predictions of various reasoners. It turned out that none of the reasoners did a perfect job of it. Even though people did tend in the majority to classify problems as ambiguous, that is insufficient evidence to opt for the most skeptical possible system (presumably, downwards reasoning without even preemption) since there may be other factors, like problem complexity, that interact with people's classifications. The concentrations of responses to problems after the indeterminate classifications are set aside yield evidence for less skeptical systems, or at least a stronger foothold from which to explore the factors that affect their classification. It is equally interesting to define reasoners which yield a closer approximation to observed behavior. Before offering a parameterized implementation for the systems discussed in the last three sections, I'll define a few more systems which come closer to fitting with the responses that people actually made to the salient graphs.

In general terms, people's behavior was in accord with transitivity for simple problems. However, a definition of complexity was not provided, although it seems that whatever definition is provided interacts with at least: the number of links in network, and the number of paths through a network. These two factors are not inter-reducible, though they tended to be related in the problems tested. Essentially, with greater complexity there was an increased trend to classify the network as indeterminate, which, interestingly, implies that people reason more closely to the way predicted by full skepticism than either extended skepticism or ideal skepticism. People rarely behaved as would have been expected if reasoning were guided by preemption. Rather than 'more-direct' links obtaining priority for one path over another, they seemed to provide conflicting information which rendered the whole problem indeterminate. Recall that this is the net effect of basing the definition of preemption on something like informational redundancy, although that definition does not capture the issues at stake (since it does still license disregarding some paths as redundant). A definition which comes closer in effect and classification would be one that assumes no link redundant, and rules out preemption as well in most cases. Shortest path reasoning (with which most theories of preemption agree on *certain* networks) was not supported at all by observation, although there was a tendency to accept direct preempting links as decisive; this supports on-path preemption without redundant links. However, that support as well also falls away with more complex problems. There did seem to be some support as well for conclusions based on counting the number of paths of each polarity, more support with respect to complex problems than was available for most of the other strategies, in fact. However, that strategy itself only holds if a rather fundamental assumption of the inheritance literature is set aside, in particular, the assumption that paths can contain negative links only as their last links. Observations suggest at least three possible alternatives, the first two seeming to fit better than the last.

Definition 24 (negative path threshold)

Let $\text{intensity}(\pi) = \begin{cases} 1 \ \textit{iff it contains at least one negative link} \\ 0 \ \textit{otherwise} \end{cases}$

Definition 25 (negative path intensity)

Let $\text{intensity}(\pi) = \textit{the number of negative links in } \pi$

Definition 26 (negative path algebra)

$$\textit{Let}\ \mathrm{intensity}(\pi) = \begin{cases} 0 & \textit{iff it contains an even number of negative links} \\ 1 & \textit{otherwise} \end{cases}$$

Definition 27 (psychologically plausible inheritance (path-counting))

1. Let π be a path.
   (a) If π is a direct link, then π is permitted.
   (b) If the degree of π is one then π is a direct link by the definition of degree, hence π is permitted.
2. Let π be a compound path of degree n. Assume that all permitted paths with degree less than n are known.
   (a) If π is a positive path, then π is permitted iff $|\ \mathit{firstnode}(\pi) \longrightarrow \omega \longrightarrow \mathit{lastnode}(\pi)| \geq |\pi'|$ for which $\mathit{firstnode}(\pi) = \mathit{firstnode}(\pi')$ and $\mathit{lastnode}(\pi) = \mathit{lastnode}(\pi')$ where for each π' $\mathit{intensity}(\pi') \geq 1$
   (b) If π is a negative path, then π is permitted iff $|\ \mathit{firstnode}(\pi) \mathrel{-\!\ominus\!\!\rightarrow} \omega \mathrel{-\!\ominus\!\!\rightarrow} \mathit{lastnode}(\pi)|$ (where $\mathit{intensity}(\pi') \geq 1) \geq |\pi'|$ for which $\mathit{firstnode}(\pi) = \mathit{firstnode}(\pi')$ and $\mathit{lastnode}(\pi) = \mathit{lastnode}(\pi')$.

4.4 Implementation

Another advantage of spelling out the definitions of various inheritance reasoners in common terms is that it facilitates implementation as well as just abstract interclassification. The first order specifications translate fairly directly into logic programs, offering a further advantage in making the relationship between the definitions and their implementation more clear. This section details a Prolog implementation of the various systems from the preceding section. The basic system was first presented by Vogel et al. (1993).

4.4.1 H90: upwards chaining, off-path preemption, restricted skepticism

Permission

A restatement of Definition 8 in Prolog is shown in Figure 4.4.1. This restatement is slightly different in structure from Definition 8; the Prolog definition is not stated explicitly in terms of increasing degree, although it relies on the relationship between the degree of a path and the degree of a subpath as stated in Proposition 1 of Chapter Two. The Prolog definitions consider only links that are relevant to the query which participate in some path sharing an endpoint with one of the queried nodes. Degree is just the length of the longest generalized path between those nodes; in the present system that concept can be accessed directly by referring to those paths directly as paths rather than indirectly via their lengths.

The base case (shown below) is the first `permitted` clause which states that all direct links in a network are permitted. The empty list in the middle position of the term for the basis indicates that no intermediate nodes lie on the path. An additional base clause to permit paths of degree one is unnecessary since the set of paths whose degree is one is a subset of the set of paths that are direct links. The second clause of the Prolog program below defines permission in the general case where π is the path from `From` to `To`. The relation `complement` is used in defining this clause; thus, the same clause stipulates the permission of both positive and negative paths. The path from `From` through `SubPath` to `Last` maps to α in the formal definition of permission, and `To` maps to z (uppercase tokens are assumed to be variables). The relations, *link* and *lastnode*, verify that $lastnode(a) \mathrel{-\!\!\ominus\!\!\rightarrow} z$ is a link contained in the network. Note that if a directly conflicting link is contained in the network, then due to the presence of the restriction `not(link(From,NotTo))`, the path from `From` to `To` is not allowed (according to condition (ii) from Definition 2). The reference to `permitted` specifies that the subpath α must itself be permitted. Finally, the relation `no_other` holds when the path specified as input through its arguments is actually a chain of links through the network which is not itself preempted. Thus, the specification `not(unpreempted)` stipulates that no conflicting, unpreempted paths exist. This is equivalent to the specification in the formal definition that is expressed: for all paths that conflict with $a \mathrel{-\!\!\ominus\!\!\rightarrow} z$, there exists some path which preempts each conflicting path. The clauses which make up the definitions of `unpreempted` and `permitted` implement Definition 8 even though a different ordering is stated on those constraints. The order of the restrictions stated in the Prolog definition is guided by efficiency considerations in limiting the search space.

`permitted(From,[],To) :-`	*Basis*
`        link(From,To).`	
`permitted(From, Path, To) :-`	π *is a compound path...*
`        nonempty(Path),`	*... from* `From` *through* `Path` *to* `To`
`        path(From, Path, To),`	*for* α
`        lastnode(SubPath, Last, Path),`	*for* π'
`        complement(To, NotTo),`	*(ii), no direct link conflicts*
`        not(link(From,NotTo)),`	*(i),* α *is permitted*
`        permitted(From, SubPath, Last),`	*(iii), just give endpoints of* π'
`        no_other(From, NotTo).`	
`no_other(From, To) :-`	*All* π' *(compound conflicts) are pre-*
`        not(unpreempted(From, _, To)).`	*empted*
`unpreempted(From,[],To) :-`	
`        link(From,To).`	*A link isn't preempted*
`unpreempted(From,Path,To) :-`	*Is a compound path* π' *not pre-*
`        nonempty(Path),`	*empted?*
`        path(From, Path, To),`	*It must be compound...*
`        lastnode(SubPath, Last, Path),`	*It must be a path* ($\pi' = \alpha' + z$)
`        positive_path(From, SubPath, Last),`	*Such that* α' *exists,*
`        allowed(From, SubPath, Last),`	*where* α' *is a positive path*
`        not(preempted(From,Path,To,By)).`	*that is allowed (ie. permitted).*
	The well-formed π' *is unpreempted*
`allowed(From, Thru, To) :-`	*if it*
`        permitted(From, Thru, To).`	*is not preempted.*

Figure 4.15 depicts a Prolog translation of the network given in Figure 4.1 (reproduced in Figure 4.16 for convenience). Links are encoded as two place relations. The polarity of a link is indicated in the second argument. A positive link $x \longrightarrow y$ is encoded as `link(x,y)`, and the negative link $x \not\longrightarrow y$ is encoded as `link(x,not(y))`. Figure 4.14 includes a Prolog session that applies the Prolog definitions given above to the network shown in Figure 4.1 and translated to Prolog in Figure 4.15. In a Prolog terminal session, user input is entered after a question

```
| ?- permitted(From, Through, To).

From = a,
Through = [],
To = b ? ;

From = a,
Through = [],
To = c ? ;

From = b,
Through = [],
To = d ? ;

From = c,
Through = [],
To = d ? ;

From = c,
Through = [],
To = not(e) ? ;

From = d,
Through = [],
To = e ? ;

From = a,
Through = [b],
To = d ? ;

From = a,
Through = [c],
To = d ? ;

From = a,
Through = [c],
To = not(e) ? ;

From = b,
Through = [d],
To = e ? ;

no
```

Figure 4.14: "Which Paths Are Permitted?"

```
link(a,b).        link(a,c).          link(b,d).
link(c,d).        link(c,not(e)).     link(d,e).
```

Figure 4.15: A Prolog Translation of Figure 4.16.

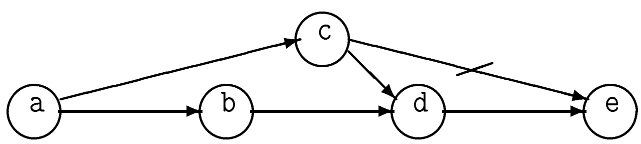


Figure 4.16: Path *abde* is Off-Path Preempted by the Link $c \not\rightarrow e$.

```
preempted(From,Path,To,Preemptor) :-
        path(From, Path, To),                % Define Pi
        lastnode(SubPath, Last, Path),
        complement(To, NotTo),
        link(Preemptor, NotTo),              % Preempting link
        relation(OtherPath, SubPath),        % On or Off path preemption
        positive_path(From, OtherPath, Last), % Define Pi'
        permitted(From,OtherPath,Last),      % Pi' is permitted
        member(Preemptor,[From|OtherPath]).  % P is on Pi' (but not Last)
```

Figure 4.17: The Definition of Off-Path Preemption in Prolog.

mark. The entry of a semicolon indicates the user's request for the interpreter to find another way to satisfy the query. The only negative paths permitted are ac/e and c/e. Six of the ten permitted paths correspond to direct links. This example shows the flexibility obtained by the Prolog specification from the logic programming paradigm in that none of the queried nodes need be specified (although, some or all nodes may be) in order to investigate the implications of the network. A more important advantage of the logic specification is its immediate mutability into other instances in the space of possible inheritance reasoners.

Off-Path Preemption

Figure 4.17 depicts a restatement the definition of off-path preemption as a Prolog relation between a path (specified in the arguments `From`, `Path`, `To`) and a preempting link from `Preemptor` to `To`. While `From`, `Path`, and `To` are all variables, it is assumed that `From` and `To` vary over nodes, and that `Path` varies over lists of nodes with positive, not negative, links implicitly connecting those nodes in the order of occurrence in a given list. The path from `From` to `Last` through `SubPath` corresponds to π in Definition 9, since the `lastnode` of `Path` is stipulated to be `Last`. The `preempted` relation can hold only if the chain of links described by the path $\mathit{From} \longrightarrow \mathit{Through} \not\longrightarrow \mathit{To}$ is actually a chain of links in the network. The relations `complement` and `link` determine whether there is a conflicting path terminating at the node `To`. The path from `From` to `Last` through `OtherPath` corresponds to π' in Definition 1—π and π' have the same first node `First` and the same last node `Last`. Off-path preemption obtains if `relation` does not impose any restrictions on its arguments. The relation `lastnode` verifies that $\mathit{lastnode}(\pi)$

```
| ?- preempted(From,Through,To,Preemptor).

From = a,
Preemptor = c,
Through = [b,d],
To = e ? ;

From = a,
Preemptor = c,
Through = [c,d],
To = e ? ;

From = c,
Preemptor = c,
Through = [d],
To = e ? ;

no
| ?-
```

Figure 4.18: "Which Paths Are Preempted?"

$= \mathit{lastnode}(\pi')$, and the invocation of `permitted` (defined above) verifies that π' is actually permitted. Finally, the call to `member` is used to determine if some node p (`Preemptor`) other than $LastNode(\pi')$ (`Last`) participates in the conflicting link.

The Prolog session reproduced in Figure 4.18, shows the application of the relation `preempted` to determine what paths between the nodes `From` and `To` of Figure 4.15 are preempted and what the preempting node is. Three preempted paths are returned: *abde*, *acde* and *cde*. All of these paths are preempted by the (negative) link from c to e—the first path is off-path preempted (since c is not on the path *abde*) while the last two paths are on-path preempted.

4.4.2 Other Systems

On-Path Preemption

Above, off-path preemption was implemented through the `relation` predicate which articulated the relationship, apart from coincidence at the endpoints, of the paths under consideration (a path and its potential preempting path). Recall that the difference between off-path and on-path preemption is that in on-path preemption one path must be a subpath of the other—the preempted path cannot contain nodes that the preempting path does not contain, although the preempted path can contain fewer nodes. This is implemented just by imposing stronger constraints on the `relation` predicate, as shown in Figure 4.19. Figure 4.20 depicts a Prolog session with the reasoner running on the same network as used to illustrate off-path preemption. In particular, note that ac/e is not a permitted

```
                                 % OtherPath and Path are the same or
relation(X, Y) :- sublist(Y,X).  % Path is redundant to OtherPath
```

Figure 4.19: The Definition of On-Path Preemption in Prolog.

```
| ?- permitted(From, Through, To).

From = a,              From = b,                 From = a,
Through = [b],         Through = [d],            Through = [c],
To = d ? ;             To = e ? ;                To = d ? ;

no
```

Figure 4.20: "Which Paths Are Permitted with On-Path Preemption?"

path. However, Figure 4.21 is one for which on-path preemption does disallow a path. The network translation is obvious; a prolog session showing the preemption of both *abcd* and *acd* is given in Figure 4.22 (permission of direct links has been edited out).

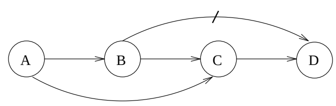


Figure 4.21: On-Path Preemption: $a \overset{b}{\not\rightsquigarrow} d$

```
| ?- permitted(From, Through, To).

From = a,
Through = [b],
To = c ? ;

From = a,
Through = [b],
To = not(d) ? ;

no
```

```
| ?- preempted(From,Through,To,By).

By = b,
From = a,
Through = [b,c],
To = d ? ;

By = b,
From = b,
Through = [c],
To = d ? ;

By = b,
From = a,
Through = [c],
To = d ? ;

no
```

Figure 4.22: "Which Paths Are Permitted with On-Path Preemption?"

No Redundancy

Obtaining on-path preemption under the assumption that no link is redundant requires the trivial re-definition of the `relation` predicate such that it requires both paths to be identical forcing on path preemption (`relation(X,X)`). I will not supply an implementation for Boutilier's (1989) system using informational redundancy rather than the usual topologically defined notion in the check on whether there are any conflicting paths that matter, nor do I supply an implementation for inheritance using informational redundancy behind its notion of permission.

Skepticism

Fully Skeptical Reasoning The difference between restricted skeptical reasoning and fully skeptical reasoning is in whether subpaths of permitted paths are required to be unambiguous. Restricted skeptical reasoners do not require subpaths to be permitted in all cases, as discussed above. The difference is encoded in the specification of the relation `allowed`, which implements part of condition (iii) from the main definition of permission. It is actually invoked in considering paths that conflict with the path under construction (`unpreempted`). Thus, the usual case requires that conflicting paths be permitted up to their penultimate node.

```
allowed(From, Thru, To) :- permitted(From, Thru, To).
```

While the fully skeptical reasoner just requires that the chain of links up to the second to last node not be preempted.

```
allowed(From, Thru, To) :- unpreempted(From, Thru, To).
```

This completes the modification required to obtain a fully skeptical reasoner. The non-preemption of α'/z must still be verified, as in the definition, because otherwise we disregard more specific information provided by a preempting link. A terminal session which reproduces the application of the revised system to the network given in Figure 4.9 (reproduced in Figure 4.23) is shown in Figure 4.24. As expected, the only permitted paths are those corresponding to each of the direct links and to the compound path ab/e.

Extended Skeptical Reasoning Extended skeptical reasoning, as introduced above, employs subpath credulity and adds reasoning with cases in order to gener-

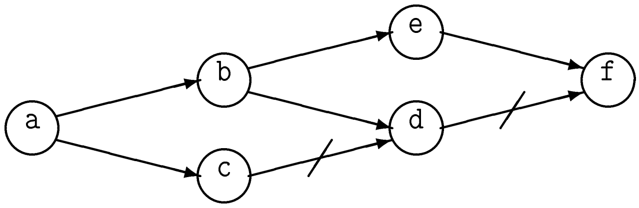


Figure 4.23: A Network with Potentially Cascaded Ambiguity.

```
| ?- permitted(From, Through, To).

From = a,
Through = [b],
To = e ;

no
```

Figure 4.24: "Fully Skeptical Reasoning"

ate conclusions allowed in the intersection of credulous expansions, without having to compute that intersection directly. That required introducing a second clause into the stipulation of what conflicting paths matter. In the Prolog implementation, that multiplies out into a second clause for permission, different solely in the last line which has a new way of classifying conflicting paths irrelevant which requires more information about the path under question. Of course, the parameterization would be considerably cleaner if this was handled by expanding the arity of `no_other`, but that would just be a trivial manipulation of the present implementation which does do the job.

```
permitted(From, Path, To) :-
        nonempty(Path),                  % Distinguish from base case
        path(From, Path, To),            % Define Alpha
        lastnode(SubPath, Last, Path),
        complement(To, NotTo),
        not(link(From,NotTo)),           % No conflicting direct link exists
        not(preempted(From, SubPath, Last,P)),  % Alpha is not preempted
        conflicts_irrelevant(From,SubPath,Last,To).
```

The new predicate, `conflicts_irrelevant`, as stipulated by the definition establishes whether a conflicting chain of links exists and requires each intermediate node to support the inference to the endpoint `To` of the path under extension

(`From` $\longrightarrow$ `Subpath` $\longrightarrow$ `Last` which is tested for extendibility to `To`), even if there would be conflicts on subpaths.

```
conflicts_irrelevant(From,Subpath,Last,To) :-
        conflict(From,Subpath,Last,BadPath,not(Bad)),
        all_permitted([Bad|BadPath],To).

all_permitted([],_).
all_permitted([Node|Path],To) :-
        permitted(Node,_,To),
        all_permitted(Path,To).

conflict(From,Subpath,Last,BadPath,Bad) :-
        path(From,BadPath,Bad),
        complement(Bad,Good),
        member(Good,[Last|Subpath]).
```

Downwards Reasoning

An initial proposal for turning the Prolog definitions into a specification of downwards reasoning might involve utilizing the reversibility of logical relations with a query like `permitted(X,Y,n)`—for some node `n`: what paths through the nodes as ordered in `Y` and beginning at `X` lead to `n`. However, this is not a correct intuition since the definition of the relation is for an upwards reasoner—with all of the variables instantiated the reasoner will attempt to prove the goal stated in the query as an upwards reasoner would. But, upwards and downwards reasoners behave differently (Touretzky et al., 1987). The alterations to the code for the basic algorithm are much the same as the changes to the definition of permission given earlier. First, it is necessary to specify path construction as in:

```
path(From,[],To) :- link(From,To).
path(From,[Mid|Through],To) :-
         link(Last,To),
         path(From,Front,Last),
         lastnode(Front,Last,[Mid|Through]).
```

And then, just supply an alternative main-case definition for permission. In fact, this one requires a bit less computation since it does not require invocation of the `lastnode` relation in order to find the last element of the path being extended. Permission of the extended path is tested against the back end of the path rather than the front end.

```
permitted(From, [First|SubPath], To) :-
        path(From,[First|SubPath], To),            % Define Alpha
        %lastnode(SubPath, Last, Path),
        complement(To, NotTo),
        not(link(From,NotTo)),          % No conflicting direct links
        permitted(First, SubPath, To),  % Alpha is permitted
        no_other(From, NotTo).
```

The versions of `no_other` and its (potentially, depending on what `no_other` invokes itself) nested call to `allowed`) can be substituted as with the forward chaining system.

4.5 Discussion

Due to the proliferation of inheritance reasoners, it is important to determine precisely how different classes of reasoners are related. This chapter outlines a declarative specification of inheritance reasoning which makes it easier to see the relationships among various systems along axes distinct from the features that define them (for instance, downwards reasoning without preemption being the most skeptical sort of system possible). The relationships among various reasoners have been illustrated by defining the various reasoners through modifications to a basic set of definitions. To my knowledge, the space of reasoners outlined here (which is larger than that considered by the "Clash of Intuitions" paper (Touretzky et al., 1987)), has never been presented within such a uniform framework. One advantage of this presentation is that it sets the stage for future work to provide rigorous analysis of exactly what equivalences exist among these systems. A second more immediate advantage is that it leads to an easier implementation of the space of reasoners by parameterizing code for the various substitutions required in the definitions.

This chapter concludes with some observations about the complexity of downwards reasoning as defined in this thesis. It is useful to first give the basic algorithm:

Input: a direct acyclic graph of positive and negative default inheritance links

Query: For all nodes x, y and through arbitrary paths π, π'..., does the network support $x \overset{\pi}{\leadsto} y$ or $x \overset{\pi}{\not\leadsto} y$.

Optimized Algorithm for Downwards Permission:

1. Let π be a path.
   - If π is a direct link, then π is permitted ($\mathcal{O}(c)$).
   - Record $x \overset{\emptyset}{\leadsto} y$ or $x \overset{\emptyset}{\not\leadsto} y$ as appropriate ($\mathcal{O}(c)$).
2. Let π be a compound path of degree n. Assume that all permitted paths with degree less than n are known.
   (a) If π is a path then it has the form $x \longrightarrow \alpha$. The path α may be positive or negative. The path π is permitted (so, record $x \overset{y\omega}{\leadsto} z$ or $x \overset{y\omega}{\not\leadsto} z$ as appropriate) iff
      i. α is recorded as permitted ($\mathit{firstnode}(\alpha) = y$, $\mathit{lastnode}(\alpha) = z$, $y \overset{\omega}{\leadsto} z$; $\mathcal{O}(n^2)$),
      ii. $x \not\longrightarrow \mathit{lastnode}(\alpha)$ is not a direct conflicting link in the net ($\mathcal{O}(c)$),
      iii. no conflicting path matters.

From the above, it is clear that the most expensive fixed step is the test on whether the subpath is itself permitted and therefore extendible into a longer path. The reason this costs $\mathcal{O}(n^2)$ (n = number of links in the network), is that the length of the list of recorded implicit links is bounded by n^2. Similarly, the number of times step two gets invoked is bounded by: $\frac{(n-1)^2+(n-1)}{2}$, which is also $\mathcal{O}(n^2)$. Thus, over the whole algorithm, the fixed steps cost $\mathcal{O}(n^4)$. The cost of determining whether conflicting paths matter is not fixed, and depends on which version is used. Consider the following:

Input: a direct acyclic graph of positive and negative default inheritance links, a list of permitted implicit links, a putative implicit link in question: either $x \overset{\pi}{\leadsto} y$ or $x \overset{\pi}{\not\leadsto} y$.

Query: are there any paths $x\omega/y$ (which conflict with $x\pi y$, if that is the input) or $x\omega y$ (which conflicts with $x\pi/y$, if that is the input) which are not preempted?

Conflicts that Matter: Restricted Skepticism No path conflicting with π matters iff

1. if π has the form, αz, and all negative paths $\pi' = \alpha'/z$, where α' is a permitted positive path ($\mathcal{O}(n^2)$) and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.
2. if π has the form, α/z, and all positive paths $\pi' = \alpha' z$, where α' is a permitted positive path ($\mathcal{O}(n^2)$) and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.

Assume that the version of preemption in place is on-path.

Input: a direct acyclic graph of positive and negative default inheritance links, a list of permitted implicit links, a putative implicit link in question: either $x \overset{\pi}{\rightsquigarrow} y$ or $x \overset{\pi}{\not\rightsquigarrow} y$.

Query: is $x\pi y$ or $x\pi/y$ (as determined by the input) preempted?

On-Path Preemption Let π and π' be positive paths, and let p and y be nodes. A positive path πy is *preempted* by a link $p \not\longrightarrow y$ or a negative path π/y is *preempted* by a link $p \longrightarrow y$ if there exists a permitted path π' such that

1. $\pi \subseteq \pi'$ ($\mathcal{O}(n)$)
2. $firstnode(\pi) = firstnode(\pi')$ ($\mathcal{O}(c)$),
3. $lastnode(\pi) = lastnode(\pi')$ ($\mathcal{O}(c)$),
4. p occurs in π' ($\mathcal{O}(n)$), and
5. $p \neq lastnode(\pi')$ ($\mathcal{O}(c)$).

Thus, on-path preemption contributes in the worst case $\mathcal{O}(n)$ to the complexity of determining whether any conflicting paths matter (using restricted skepticism), which does not overshadow the cost of determining subpath permission. Since determining whether conflicting paths matter iterates over known permitted paths, and that is bounded by $\mathcal{O}(n^2)$, the overall cost of determining whether conflicting paths matter is $\mathcal{O}(n^4)$. This means that for downwards restricted skeptical reasoning with on-path preemption, the time complexity is $\mathcal{O}(n^6)$. Note that for this worst case analysis it makes no difference to substitute off-path preemption.

However, utilizing ambiguity cascading skepticism, there is a slight improvement to $\mathcal{O}(n^5)$.

Input: a direct acyclic graph of positive and negative default inheritance links, a list of permitted implicit links, a putative implicit link in question: either $x \overset{\pi}{\leadsto} y$ or $x \overset{\pi}{\not\leadsto} y$.

Query: are there any paths $x\omega/y$ (which conflict with $x\pi y$, if that is the input) or $x\omega y$ (which conflicts with $x\pi/y$, if that is the input) which are not preempted?

Conflicts that Matter: Subpath Credulity No path conflicting with π matters iff

1. if π has the form, αz, and all negative paths $\pi' = \alpha'/z$, where α' is a positive path ($\mathcal{O}(n)$) and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.
2. if π has the form, α/z, and all positive paths $\pi' = \alpha' z$, where α' is a positive path ($\mathcal{O}(n)$) and $firstnode(\alpha) = firstnode(\alpha')$, are preempted.

Tighter bounds are almost certainly possible, but the present results are sufficient to satisfy the intuition that downwards reasoning need not be intractable if chaining is defined symmetrically to the forward chaining case. In particular, if H90 inferences, a restricted skeptical, off-path preempting, upwards chaining reasoner are polynomial computable, then so should a downwards version. The importance of this is to be rated in light of the intractability of downwards reasoning reported by Selman and Levesque (1989, 1993): however it should be emphasized that a polynomial algorithm for downwards inheritance as defined here does not contradict their results; rather it emphasizes the asymmetry they assume between the definitions of upwards and downwards reasoning. The downward reasoning they consider is actually double chaining. It also turns out that downwards fully skeptical reasoners are more efficient than the restricted skeptical version. This makes it worth examining whether the same is true for the upwards case. I will not analyze that possibility in this thesis; but just point out again that the declarative framework in which the reasoners are defined here makes it quite possible for such modular analyses to be performed.

Chapter 5

Channel Theoretic Semantics for Inheritance Reasoning

5.1 Introduction

Inheritance networks provide a formal tool for reasoning about classifications of individuals and concepts as an alternative to first order logic (FOL). Inheritance reasoning is analogous to first order deduction, but there are important differences. One major difference is that the dominant approaches to path based inheritance do not admit quantification. Additionally, with respect to inferences derivable from inconsistent sentences, inheritance reasoners localize inconsistency and do not sanction arbitrary conclusions. Traditionally, relationships between individuals and classes have been stated in terms of the relations *is-a* and *is-not-a*, but other relations are possible (Brachman, 1983; Touretzky, 1986). The relations may be strict or defeasible and networks of links that represent those relations may contain both sorts. By reasoning over such a network, one may determine the properties possessed by an individual. For a given network, there are numerous reasoning methods. Horty has observed that there are over 72 different path-based inheritance reasoners for an inheritance network containing *is-a* and *is-not-a* relations (Horty, 1989). These reasoners differ in the manner in which they treat conflicting information. While Touretzky (1986) provided a lattice based

[0]Thanks to Jon Barwise and Lawrence Cavedon for extensive comments on material contained in this chapter.

semantics for his first exploration of the area, it has been pointed out that a great deal of the literature focuses on proof theory rather than general semantics of inheritance reasoners (e.g. Boutilier, 1989). Semantics have been provided by a number of researchers for individual reasoning systems (Delgrande, 1990; Boutilier, 1989; Bacchus, 1989), but no formal semantics has been outlined which provides a general enough framework to capture a large class of reasoning systems. This chapter hopes to lay the groundwork for such a semantics for the class of inheritance reasoners defined by Touretzky et al. (Touretzky, 1986; Touretzky et al., 1987; Horty & Thomason, 1988; Horty et al., 1990) by taking advantage of the tools for talking about natural constraints provided by situation theory and channel theory. (Barwise, 1989, 1991, 1993; Barwise & Seligman, 1994).

The axes of variation in the class of reasoners considered are upwards and downwards, skeptical and credulous reasoning when faced with on-path or off-path preemption. Implications of choices with respect to these axes are outlined by Touretzky et al. (1987) and in Chapters Two and Four of this thesis. This chapter intends to offer a channel-theoretic semantics for H90 (a system which currently lacks semantics) and to appeal to the parameterization in the proof theory to define clear semantics for the other reasoners in the class. The next section begins the groundwork for providing a semantics to H90 by outlining the situation theoretic tools which will be used to model inheritance. A channel theoretic model of nonclassical feature structures (related to the one presented by Vogel and Cooper (1994)) is provided as a preliminary example. I argue that channels provide an ideal level of description for providing a semantics to default inheritance reasoning. The central idea of channel theory is that the connection between tokens is a real object unto itself, an information channel, and supports information flow (constraints). The idea is applied here by using channel types as the interpretation of default inheritance links, and identifying the semantics of various proof theories for path-based inheritance (which derive conclusions from networks from the paths that are permitted through them) with corresponding restrictions on the composition of atomic channels (the denotations of inheritance links) into composite channels (the denotations of paths). This technique is demonstrated by providing a model of skeptical inheritance embodied by the H90 system, as well a number of other proof theoretic variations. The specific channel theoretic tools specialized for the semantics of inheritance were initially presented by Vogel (1992), but the material has developed considerably since then.

5.2 Modeling Tools

Developing semantic models for inheritance reasoning is no easy task as is easily seen from the lack of a general semantics for inheritance. It makes good sense to start out with a theory that offers a great number of modeling tools, and situation theory seems to have quite a lot to offer in that regard. In particular, the theory of constraints from situation theory (Barwise, 1993) augmented with the theory of information channels from channel theory (Seligman, 1990; Barwise & Seligman, 1994; Barwise, 1993; Barwise & Seligman, 1993; Seligman & Barwise, 1993) provides a promising framework.

The components are sites (situations), types, and channels, with relations among the three kinds of objects. $\{T^1, T^2, \ldots\}$ are types; $\{s^1, s^2, \ldots\}$ are sites; $\{c^1, c^2, \ldots\}$ are channels. Let $\{C^1, C^2, \ldots\}$ be types associated with channels. *Signaling relations* pair sites with other sites, relative to a channel ($s^1 \stackrel{c^1}{\mapsto} s^2$). *Indicating relations* pair types with other types ($T^1 \Rightarrow T^2$). The *of-type* relation pairs sites with types ($s^1 \models T^1$), channels with channel types ($c^1 \models C^1$), and channel types with indicating relations ($C^1 \models (T^1 \Rightarrow T^2)$). The of-type relation is sometimes written as a colon (:). It is sometimes easiest to refer to sets of signaling relation 3-tuples in terms of channel types rather than channel tokens ($s^1 \stackrel{C^1}{\mapsto} s^2$).

Definition 28 *Channel Types. Let $\mathcal{CT}$ be a total function $(c \times C)$ such that $\mathcal{CT}(c') = C'$ iff $c' \models C'$. It is assumed that if $\mathcal{CT}(c') = C'$ and $\mathcal{CT}(c') = C''$ then $C' = C''$.*

I is a function that assigns situation types to nodes in an inheritance network. A model is a tuple: $\langle S, T, C, I, \models, \mapsto \rangle$. Restrictions on the elements of this tuple which obtain models of various inheritance reasoners will be discussed after some essential definitions are spelled out.

It should be made clear that the mathematics of channel theory is not fully worked out. A series of recent papers has established the potential applicability of the theory to a range of philosophical and mathematical problems (Seligman, 1990;

Barwise & Seligman, 1994; Barwise, 1993; Barwise & Seligman, 1993; Seligman & Barwise, 1993) but like situation theory itself, there is no universally accepted notation nor set of axioms that detail the necessary and sufficient conditions for a channel theoretic model to be a channel-theoretic model. Some of the definitions that I will use do not correspond exactly with the definitions provided to the concepts in other works; however, the definitions do intuitively fit these concepts and their present invocation.

The basic idea of this chapter is that an inheritance network translates fairly directly into an interpretation involving indicating relations. Each default link in an inheritance network is understood as a constraint in situation theoretic terms. Constraints are interpreted as indicating relations on types of situations and are supported by information channels. Barwise argues that an indicating relationship models constraints that hold between, for instance, the height of columns in thermometers and the temperature of the surrounding environments, while signaling relationships model the connection between a particular thermometer in a particular environment, through the natural regularity that connects thermometers and temperatures. Constraints that hold in general may or may not hold in particular. The key to reasoning is composition of the relations named above, but this is only sensible when certain conditions hold with respect to the signaling relations, the actual connections, underneath the indicating relations.

5.2.1 Tools from Channel Theory

First, it is useful to provide some definitions to constrain models. Essentially, these define aspects of unimpeded information flow.

Definition 29 *Given a signaling relation* $s_1 \overset{c}{\mapsto} s_2$*,* s_1 *is a* signal *for* s_2 *relative to* c *and* s_2 *is a* target *for* s_1 *relative to* c *(Barwise, 1993).*

A signaling relation is a three-place relation between signal and target tokens and channel tokens. A channel, in this sense corresponds to that of Barwise (1993) and approximately to the sense of *link*[1] used by Barwise and Seligman (1993).

[1] This use of *link* is again different from the sense in this paper—that of a link in an inheritance hierarchy. For Barwise and Seligman (1993), *link* denotes certain kinds of connections between classification systems.

Definition 30 $C \models T_1 \Rightarrow T_2$ *iff for all* $c \models C$, $c \models T_1 \Rightarrow T_2$.

Definition 31 *If* $C \models T_1 \Rightarrow T_2$ *then* T_1 indicates T_2 *relative to* C *(Barwise, 1993).*

A basic contribution of channel theory is that information channels support constraints. Channels are essentially conduits of information such that if a particular channel is of a certain channel type, then it supports a constraint, and links tokens in such a way that if the signal supports the type in the antecedent of the constraint then the target is classifiable in terms of the consequent of the constraint. Thus, channels license classification of target sites in terms of the consequent types in the constraints supported by the channel type.

Definition 32 *A channel type* $C \models T_1 \Rightarrow T_2$, *is* informative *iff there exists a channel* $c \models C$ *and* s_1 *and* s_2 *such that* $s_1 \overset{c}{\mapsto} s_2$ *with* $s_1 \models T_1$ *and* $s_2 \models T_2$.

Definition 33 *If* C *is an informative channel type supporting the constraint* $T_1 \Rightarrow T_2$, *then for all sites* s, s' *such that* $s \overset{c}{\mapsto} s'$, $s' \overset{C}{\models} T_2$.

The modeling power of channel theory comes from the potential difference between the types supported by a token site ($s' \models T_2$) and the types that a token site can be classified in terms of due to the constraints that site is connected to through information channels ($s' \overset{C}{\models} T_2$).

Definition 34 *A channel type* $C \models T_1 \Rightarrow T_2$, *is* sound *iff for all* $c \models C$ *and for all tokens* s_1 *and* s_2 *if* $s_1 \overset{c}{\mapsto} s_2$ *and* $s_1 \models T_1$ *then* $s_2 \models T_2$.

A channel type is sound if all the channels of that type support the classification of targets with types that agree with the types supported by the targets on their own.

Definition 35 *A constraint* $T_1 \Rightarrow T_2$ *is* sound *iff it is the type of some sound channel, and the constraint is* informative *iff it is the type of some informative channel.*

Seligman and Barwise (1993) characterize channel theory as a theory of information flow. Classification offers the basic unit of information flow: some token is of some type. Channel theory provides a formal means for discussing constraints because it allows description of some situation being of some type as carrying information that some other situation is of a particular type by virtue of the connectedness of the two tokens. I distinguish between sound and informative constraints. Both are types of regularity, but informative constraints are the weaker of the two. A sound constraint is one that classifies a channel whose signal/target pairs in all of the signaling relations it participates in are classified by types consistent with those predicted by the constraint. A regularity, embodied by a constraint, is informative if there is at least one instance of it grounded in a signal/target pair of the predicted types. This condition is stronger than if the constraint did not exist but two situations (corresponding to the signal/target pair) did happen to support the same types.

Defaults are interpreted as informative constraints. Necessarily exceptionless statements (squares have four sides) are interpreted with sound constraints over reflexive channels. Defaults are generally unsound. Note that a site can be classified by inconsistent types. Such a site can still be part of an informative channel, but not of a sound channel. Channel theory provides an idea of information flow; this model will take advantage of the information that some situation which supports a particular type can carry about another situation. When types are fixed as sentences of first order logic, sites are worlds, and channels are reflexive, then channel theory provides a straightforward semantics for classical logic (Barwise, 1993). Channel theory has also been used to provide an analysis of Default Logic (Cavedon, 1993).

Given a particular sentence, *Birds fly* represented with a link *Birds* $\longrightarrow$ *Fliers*, the sentence is interpreted using channels in the following way: there is an indicating relation between situations of a type *Bird* and of the type *flier*; underneath this indicating relation is a signaling relation:

$$\begin{array}{ccc} Birds & \Longrightarrow & Fliers \\ \rotatebox{90}{$\Vvdash$} & & \rotatebox{90}{$\Vvdash$} \\ s^1 & \overset{c}{\mapsto} & s^2 \end{array}$$

The signaling relation is a connection which is of a channel type that supports the constraint $Birds \Rightarrow Fliers$ and links two situations, one that supports the fact that a bird is present and another that supports the fact that a flier is present.

Note that the target site could also support the fact that a flier is not present.

$$\begin{array}{cccc} Birds & \Longrightarrow & Fliers & \\ \Downarrow & & C\Downarrow & \\ s^1 & \overset{c}{\mapsto} & s^2 & \models \neg Fliers \end{array}$$

If I watch a bird eating breadcrumbs in a park, then I am in a situation in which a flier is present; through the information channel, the bird is classifiable as a flier. However, I am also in a situation in which a flier is not present, as the bird is eating breadcrumbs and not flying. The constraint is informative because there are situations which exist in the signaling relation that support the right types, but it is not sound since there are some situations in which birds are present which are not also situations in which fliers are present.

It is perhaps not customary to assume that situations can be inconsistent, however, to provide a proper interpretation to inheritance reasoning it seems essential to allow this, at least in the weak form of inconsistency advocated here. This is because all of the reasoners in the family considered here adopt the position that all direct links in a network are supported as permitted paths by the reasoners. This means that networks containing directly conflicting links contain inconsistent information. An advantage of inheritance reasoning over other systems of logic is that the impacts of inconsistency on reasoning are localized. Since people arguably sometimes maintain inconsistent beliefs even after becoming aware of the inconsistency, it is important to be able to maintain inconsistent situations as potential parts of models (cf. Braisby, Franks, & Hampton, 1994). Weak inconsistency is allowed in this model in that it is not required that there be agreement between the direct typing of sites $s \models T$ and the typing of a site that results from its participation in some channel $s \overset{C}{\models} \neg T$. That is, a certain constraint may convey information about some token that is inconsistent with the way the token actually is.

The network depicted in Figure 5.1 is more complex than the case of flying birds since it represents two contradictory defaults. If both of the constraints that interpret those defaults are informative, then there can be situations which support the type B which signal other situations that support both the type D and the type $\neg D$, but necessarily there are some situations that carry D and some that carry $\neg D$ (those situations are not necessarily the same ones). However, it

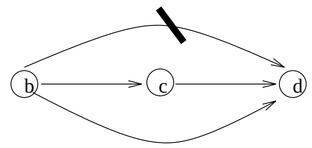


Figure 5.1: An Inconsistent Network.

is possible to assign an interpretation to networks such that we would want to assert that the model includes inconsistent sites. Let b and d both be types of linguistic information.[2] Let the sites be utterance situations, and let the channels be the ones pertaining to the constraints of English grammar. Now, the utterance of the sentence, "The examples is convincing," conveys conflicting information about the number of the subject. The NP "The examples" on its own suggests that the subject of the sentence is plural, while the VP indicates that the subject should be singular. Here we have an utterance situation in which the subject is both singular and plural, by the constraints of English grammar. However, the inconsistency is localized, since other semantic contents of the sentence can be processed without difficulty.

Definition 36 *A full model of a network is informative if all of its constraints are informative, and it is sound if all of its constraints are sound.*

5.2.2 Defeasibility

Information does not always flow as it seems it should. The idea of channel theory is that channels are still able to model regularity in the world, even when information flow does not correspond to that regularity in particular circumstances. The following definitions are introduced to model the conditions that arise with the interpretation of defaults.

Definition 37 *A channel type* $C \models T_1 \Rightarrow T_2$ *has a* pseudosignal s_1 *iff* $\exists\ c \models C$ *and* $s_1 \models T_1$ *but there is no* s_2 *such that* $s_1 \overset{c}{\mapsto} s_2$.

Definition 38 *A constraint* $T_1 \Rightarrow T_2$ *has a* pseudosignal s_1 *iff the constraint is the type of a channel* c *which has* s_1 *as a pseudosignal.*

[2] Ignore the node c for the present example.

Given a constraint on situation types, a pseudosignal to that constraint can be understood as an event of the type antecedent to the constraint which is not connected by that constraint to any event which is of the consequent type. Some other event out there may be of the consequent type, but the pseudosignal is not connected to it. For example, Opus the penguin is a pseudosignal for the constraint that *Birds fly*, even if on some particular occasion Opus is flying on a Pan-Am jet. Though the second situation supports the right type, it is not connected via a channel of the type, *Birds fly.*

Proposition 2 *A sound channel can have pseudosignals.*

> **Proof:** Let $M = \langle S, T, C, \models, \Rightarrow, \mapsto \rangle$; $S = \{s_1\}$, $T = \{T_1, T_2\}$, $C = \{c : T_1 \Rightarrow T_2\}$, $\models = \{\langle s_1, T_1\rangle\}$, $\Rightarrow = \{\langle T_1, T_2\rangle\}$, $\mapsto = \{\}$. In this model, s_1 is a pseudosignal to c. But, c is also a sound channel by vacuous satisfaction of Definition 34. $\square$

Intuitively, a sound channel has a pseudosignal in any site that is classified by a type that predicts that the site should be a signal, but which is indeed not a signal to any target through the channel.

Proposition 3 *An informative channel can have a pseudosignal.*

> **Proof:** Let $M = \langle S, T, C, \models, \Rightarrow, \mapsto \rangle$; $S = \{s_1, s_2, s_3\}$, $T = \{T_1, T_2\}$, $C = \{c : T_1 \Rightarrow T_2\}$, $\models = \{\langle s_1, T_1\rangle, \langle s_2, T_2\rangle, \langle s_2, \neg T_2\rangle\ \langle s_3, T_1\rangle, \}$, $\Rightarrow = \{\langle T_1, T_2\rangle\}$, $\mapsto = \{\langle s_1, s_2, c\rangle\}$. In this model, c is informative since it participates in a signaling relationship that is grounded by sites in the model ($s_1 \overset{c}{\mapsto} s_2$, $s_1 \models T_1$ and $s_2 \models T_2$). But, also, $s_3 \models T_1$ though no other site is signaled by s_3; thus, s_3 is a pseudosignal for c. $\square$

An informative constraint can have a pseudosignal for the same reasons that sound constraints can. In the above proof it is shown that the channel can be informative because of s_1 and s_2 while having a pseudosignal in s_3.

An example of a pseudosignal is given in (5.1).

(5.1)
$$\begin{array}{ccc} T^1 & \Longrightarrow & T^2 \\ \perp\!\!\!\perp\!\!\!\perp & & \\ s^1 & & \end{array}$$

The nature of defaults is that they have exceptions.

Definition 39 *A channel type $C \models T_1 \Rightarrow T_2$, has an* exception s_1 *iff there exists $c \models C$ and there are tokens s_1 and s_2 such that $s_1 \overset{c}{\mapsto} s_2$ and $s_1 \models T_1$ but $s_2 \not\models T_2$.*

An exception is closely related to a pseudosignal in that it is a point of breakdown in classifications. In an exception to a constraint, there is an event that is of the type antecedent to the constraint, and it is connected by a channel that supports the type of the constraint to another event, but that connected event fails to be of the consequent type. For example, Jonathan the seagull that I watched eat breadcrumbs in the park is a signal for the constraint that *Birds fly*, and can be connected via a channel which supports that constraint to a situation in which Jonathan is eating, one which does not support the consequent type of *flight*. Examples of two sorts of exceptions are shown in (5.2 & 5.3). In both cases, if the channel type C is otherwise informative then $s^2 \overset{C}{\models} T^2$, even though $s^2 \not\models T^2$. Since any target of a channel is classifiable in terms of the consequent type in the constraint that the channel type supports, it is not necessary to indicate that classification (e.g. $s^2 \overset{C}{\models} T^2$) in diagrams like (5.2 & 5.3).

(5.2)
$$\begin{array}{ccc} T^1 & \Longrightarrow & T^2 \\ \rotatebox{90}{$\models$} & & \\ s^1 & \overset{c}{\mapsto} & s^2 \end{array}$$

(5.3)
$$\begin{array}{cccc} T^1 & \Longrightarrow & T^2 & \\ \rotatebox{90}{$\models$} & & & \\ s^1 & \overset{c}{\mapsto} & s^2 & \models \neg T^2 \end{array}$$

Proposition 4

1. *A signal can be an exception.*
2. *A sound channel cannot have exceptions.*
3. *An informative channel can have exceptions.*

A key feature of default inheritance reasoning is that the meaning of certain links is mitigated by the surrounding network. For instance, exceptions to certain defaults arise from conflicting defaults. The next definitions articulate important relations that can exist between a constraint and other accepted constraints.

Definition 40 *A channel type $C \models T_i \Rightarrow T_j$, has a* dual signal s_k *iff C is informative and there exists an informative channel type C' which supports the type $T_k \Rightarrow \neg T_j$ and $s_k \models T_k$.*

Definition 41 *A constraint $T_i \Rightarrow T_j$ has a* dual constraint *iff the constraint is the type of a channel c which has s_k as a dual signal.*

Dual constraints can be understood in classical logic terms as implications which can both be true at the same time only when one of them is vacuously true. For an informative constraint to have a dual is for there to be another informative constraint that carries the opposite information, even though it may be information about different source and target tokens. A graphic depiction of classification relations involved in a pair of dual channels is given in Figure 5.2.

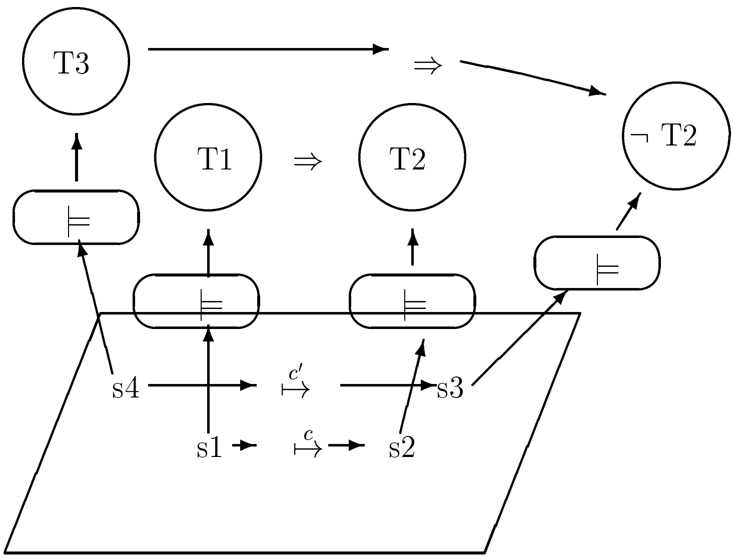


Figure 5.2: Sites $s1$ and s_4 are Dual Signals.

Proposition 5

1. *A sound channel can have a dual signal.*
2. *An informative channel can have a dual signal.*

A condition stronger than the existence of dual signals is the existence of antisignals. A pair of antisignals is a pair of dual signals that share targets.

Definition 42 *A channel $c \models T_i \Rightarrow T_j$, has an* antisignal s_k *iff s_k is a dual signal for C through C' and any target relative to C is also a target relative to C'.*

Definition 43 *A constraint $T_i \Rightarrow T_j$ has an antisignal s_k iff the constraint is the type of a channel c which has s_k as an antisignal.*

An antisignal to a constraint is something which conveys contradictory information about target tokens. An antisignal is the source of information which may flow through a completely different sort of channel that some token has more than one and conflicting types. An example of a set of classifications that establishes an antisignal is given in (5.4). Let $c^1, c^2, c^3 \models C$ and $c^4, c^5, c^6 \models C'$; thus, $C \models T^0 \Rightarrow T'$ and $C' \models T^1 \Rightarrow \neg T'$. Both C and C' are informative: C is because $s' \models T'$, and C' is because $s'' \models \neg T'$, so s^4, s^5, s^6 are dual signals to C and s^1, s^2, s^3 are dual signals to C'. Since s', s'' and s''' are targets to both C and C', s^4, s^5, s^6 are antisignals to C and s^1, s^2, s^3 are antisignals to C'. Note that $s', s'', s''' \overset{C}{\models} T'$ and $s', s'', s''' \overset{C'}{\models} \neg T'$.

$$
(5.4) \qquad
\begin{array}{|cccc|cccc|}
\hline
T^0 & \Longrightarrow & T' & & T^1 & \Longrightarrow & \neg T' & \\
\rotatebox[origin=c]{-90}{$\models$} & & \rotatebox[origin=c]{-90}{$\models$} & & \rotatebox[origin=c]{-90}{$\models$} & & & \\
s^1 & \overset{c^1}{\mapsto} & s' & & s^4 & \overset{c^4}{\mapsto} & s' & \models T' \\
 & & & & & & & \\
T^0 & \Longrightarrow & T' & & T^1 & \Longrightarrow & \neg T' & \\
\rotatebox[origin=c]{-90}{$\models$} & & & & \rotatebox[origin=c]{-90}{$\models$} & & \rotatebox[origin=c]{-90}{$\models$} & \\
s^2 & \overset{c^2}{\mapsto} & s'' & \models \neg T' & s^5 & \overset{c^5}{\mapsto} & s'' & \\
 & & & & & & & \\
T^0 & \Longrightarrow & T' & & T^1 & \Longrightarrow & \neg T' & \\
\rotatebox[origin=c]{-90}{$\models$} & & & & \rotatebox[origin=c]{-90}{$\models$} & & & \\
s^3 & \overset{c^3}{\mapsto} & s''' & & s^6 & \overset{c^6}{\mapsto} & s''' & \\
\hline
\end{array}
$$

Proposition 6

1. *A sound channel cannot have an antisignal.*

2. *An informative channel can have an antisignal.*

A special case of interpretation of conflicting defaults arises when it turns out that the same object is classified in inconsistent ways by these defaults. In channel theoretic terms this works out with a site supporting conflicting types through opposing channels. The two channels are called antichannels. Antichannels are the channels employed by signals that are their own antisignals.

Definition 44 *An informative channel* $C \models T_1 \Rightarrow T_2$ *has an* antichannel C' *iff* $C' \models T_1 \Rightarrow \neg T_2$ *is informative and all* $s_i \models T_1$ *that are antisignals for* C' *are also signals for* C.

Proposition 7

- *A sound channel cannot have an antichannel.*
- *An informative channel can have an antichannel.*

An example configuration of types and signals that constitutes an antichannel is given in (5.5). Let $c^1, c^2 \models C$ and $c^3, c^4 \models C'$. So, $C \models T \Rightarrow T'$ and $C' \models T \Rightarrow \neg T'$. Both C and C' are informative. The signals to C' are dual signals to C and vice versa. Since s', s'' and are targets to both C and C', the signals to C' are also antisignals to C and also vice versa. However, the signals to C are the signals to C'. Thus, C and C' are antichannels. As in (5.4), $s', s'' \overset{C}{\models} T'$ and $s', s'' \overset{C'}{\models} \neg T'$.

(5.5)
$$
\begin{array}{|cccc|cccc|}
\hline
T & \Longrightarrow & T' & & T & \Longrightarrow & \neg T' & \\
\rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & & & \\
s^1 & \overset{c^1}{\mapsto} & s' & & s^1 & \overset{c^3}{\mapsto} & s' & \models T' \\
 & & & & & & & \\
T & \Longrightarrow & T' & & T & \Longrightarrow & \neg T' & \\
\rotatebox{90}{$\models$} & & & & \rotatebox{90}{$\models$} & & \rotatebox{90}{$\models$} & \\
s^2 & \overset{c^2}{\mapsto} & s'' & \models \neg T' & s^2 & \overset{c^4}{\mapsto} & s'' & \\
\hline
\end{array}
$$

Definition 45 *A model is* strict *if it is informative and if none of its channels have antisignals or antichannels, and it is* defeasible *otherwise.*

5.2.3 Inference

A natural channel theoretic interpretation of inference in classical nonmodal propositional logic requires only a single token. Types correspond to sentences in the logic. Conclusions from inferences can be looked up in the semantics in the types assigned to a token. The sentence $A \supset B$ is true iff there is a constraint $A \Rightarrow B$ that is the type of an identity channel on the single token. If the channel is sound, then if the token is of type A then the token will also be classified as B. This is essentially the idea of a *redescription* channel.

Inference in inheritance networks with defaults is not classically valid. The channel theoretic model that is being sketched in this chapter offers a way to explain both why it is not valid and why it is reasonable. As has been said, inheritance links are interpreted as informative constraints, and the existence of a channel type that supports a constraint is sufficient to classify target tokens relative to the channel type with the type that is consequent to the constraint. The channels that underlie the constraints generally link distinct tokens. Information that some token is one way carries information about another token being some other way. Of course, token identity is possible, but the framework admits much more into its descriptive auspices. On the other hand, inference is performed using inheritance as an efficient way of calculating properties of some object at hand,[3] and other objects whose properties are conveyed by constraints are not entirely relevant to an inference, although the properties possessed by those objects are relevant. Inference in default inheritance networks is aptly interpreted by the projection of types related to tokens via constraints (the types of targets) upon signaling tokens, analogously to the way an informative channel licenses a particular classification of its target tokens. Thus, inference using the constraint *Birds fly* when applied to the token *Opus*, which is a bird (and in the absence of other constraints), leads to the classification of *Opus* as a *Flier*. Consider (5.6), with the assumption that s^1 is a situation that contains *Opus* and that s^2 is a situation that contains *Opus* as well as flying birds; $s^2 \overset{C}{\models}$ *Fliers* as well as $s^2 \models$ *Fliers*, but also $s^1 \overset{C^{-1}}{\models}$ *Fliers*.[4] This is distinct from a redescription channel

[3] Assuredly, there are plenty of inefficient inheritance reasoners available, but the system of Horty et al. (1990) is polynomial.

[4] Here C^{-1} is used informally to denote the inverse flow of information through a channel—this

in which s^1 and s^2 would actually be the same token (5.7). This is true even though the constraint really means that some bird-situations are connected to certain flier-situations. The room for error in projecting classifications accounts for the classical invalidity of default inheritance, but its reasonableness—apart from compact representation—is also arguable from the similarity to the model of inference in the classical case. The purpose of inference using defaults is to avoid having to look up information in a table. In this case it would mean looking up the *Opus* token in that related classification to see if *Opus* is flying in there, and similarly for any other object that one could try to reason about.

$$
\begin{array}{ccc}
Birds & \Longrightarrow & Fliers \\
\perp\!\!\!\perp & & C\perp\!\!\!\perp \\
s^1 & \overset{c}{\mapsto} & s^2
\end{array}
\tag{5.6}
$$

$$
\begin{array}{ccc}
{}^{\circ}Celsius & \Longrightarrow & {}^{\circ}Fahrenheit \\
\perp\!\!\!\perp & & \perp\!\!\!\perp \\
s^1 & \overset{c}{\mapsto} & s^1
\end{array}
\tag{5.7}
$$

Definition 46 *Let $\chi = \langle S, T, C, \rightarrow, \Rightarrow, \models\rangle$ be a system of constraints on a classification domain. The* leftward projection *of an informative constraint $T_1 \Rightarrow T_2$ down a channel c of type C such that $C \models T_1 \Rightarrow T_2$ upon a specific token s_o that is a signal for the constraint yields a new classification and system of constraints, χ' such that:*

1. $S' = S, T' = T, C' = C, \Rightarrow' = \Rightarrow$
2. $\rightarrow' = \rightarrow$ *except for those elements of the signaling relation that use the channel c. Replace each triple $\langle s_o, c, s_i\rangle$ in $\rightarrow$ with $\langle s_o, c, s_o\rangle$ in $\rightarrow'$.*
3. $\models' = \models$ *except for those elements of the of-type relation that use sites and types salient to the constraint projected upon the source. Specifically, replace each $\langle s_i, T_j\rangle$ such that $\langle c, T_h \Rightarrow T_j\rangle$ is in $\models$ and $\langle s_o, c, s_i\rangle$ is in $\rightarrow$, with $\langle s_o, T_j\rangle$.*

The leftward projection given in Definition 46 formalizes the notion described above (see Seligman and Barwise (1993) for alternative forms of projections). Other forms of projection are possible, but this particular form is closer to intuitions at stake in modeling default inheritance reasoning. It is focused upon a

---

is an abbreviation for a leftward projection which will be defined presently.

specific token s_o about which reasoning is performed. All of the types, tokens, channels, and constraints from the original classification remain in the projection, but the signaling and of-type relations change. The change in the of-type relation is the essence of the projection: types supported by a target relative to the channel the constraint is projected along are supported by the source instead. Since the constraint and channel remain, the leftward projection also redirects the signaling relation so that the former target is no longer signaled, but the source which supports the target's types is signaled.

Definition 46 implements a psychologically plausible model of inference with one important exception: it transfers inconsistency about target tokens to the projection. Although I have argued the necessity of modeling at least weakly inconsistent situations, a model of psychologically plausible human inference must account for the overwhelming desire to rationalize inconsistencies. Channel theory provides a convenient mechanism for doing this. In the most extreme case, suppose there are two directly conflicting constraints that have the same source and target tokens, and each of the constraints is the type of a distinct channel. As channel theory provides a theory of information flow, it allows the description of one type "flowing" by virtue of one channel, and the conflicting information by virtue of the other. However, at the target token, there are simply contradictory types. Inference which draws attention to the contradiction must admit a way of rationalizing it: $s_t \models T$ *because* of c and $s_t \models \neg T$ because of c'. Effectively, the type of the relevant channel is invoked to explain the source of the information. Assuming a model for the type of a channel, it is possible to incorporate rationalization into a projection upon a source.

Definition 47 *Let* $\chi = \langle S, T, C, \rightarrow, \Rightarrow, \models \rangle$ *be a system of constraints on a classification domain. The* rationalized leftward projection *of an informative constraint* $T_1 \Rightarrow T_2$ *down a channel* c *such that* $c \models C$ *and* $C \models T_1 \Rightarrow T_2$ *upon a specific token* s_o *that is a signal for the constraint yields a new classification and system of constraints,* χ' *such that:*

1. $S' = S, C' = C, \Rightarrow' = \Rightarrow$
2. $\rightarrow' = \rightarrow$ *except for those elements of the signaling relation that use the channel* c*. Replace each triple* $\langle s_o, c, s_i \rangle$ *in* $\rightarrow$ *with* $\langle s_o, c, s_o \rangle$ *in* $\rightarrow'$.

3. $\models' = \models$ *except for those elements of the of-type relation that use sites and types salient to the constraint projected upon the source. Specifically, replace each* $\langle s_i, T_j \rangle$ *such that* $\langle C, T_h \Rightarrow T_j \rangle$ *is in* $\models$ *where* $\langle s_o, c, s_i \rangle$ *is in* $\rightarrow$*, replace* $\langle s_i, T_j \rangle$ *with* $\langle s_o, T_j \wedge C \rangle$.
4. $T' = T \cup \{C \wedge T_j\}$ *for* T_j *such that* $C \models T_h \Rightarrow T_j$*,* $s_o \overset{c}{\mapsto} s_i$ *and* $s_i \models T_j$.

An important property of the rationalized leftward projected classification is that it is a coherent classification iff the antecedent classification is coherent and the channel does not support constraints that would be inconsistent if satisfied by the same sites. The latter property means that if the channel c which is projected along supports both $T_3 \Rightarrow T$ and $T_4 \Rightarrow \neg T$ then the rationalized leftward projection will contain an inconsistency (subject to the conditions on the token projected upon). Because, the information about which channel supplies the projected type is part of the type that actually gets assigned to source sites, inconsistent typing is never projected from targets. Inconsistencies that are projected in the leftward projection simpliciter are rationalized away by invoking the 'types' of the channels involved. Instead of $s \models T$ and $s \models \neg T$, there is $s \models T \wedge C'$ and $s \models \neg T \wedge C''$. This classification is consistent, but the points of inconsistency are derivable from the rationalization.

A semantics for default inheritance reasoning must predict exactly which types should get projected upon source tokens by virtue of the constraints they underlie. Clearly, since inference is conducted over long paths of links and not just single links, the composition of channels is important to the definition of a model for a specific inheritance reasoner. Section 5.3.2 will define the conditions on channel composition that define the closure of a set of informative constraints interpreting an inheritance network. The classification obtained by projecting types along the channels in the closure will constitute a model for default inheritance reasoning.

5.2.4 An Example Application: Feature Structures

Before proceeding to define the additional channel theoretic concepts requisite to the semantics of inheritance reasoning, it is useful to put the tools to an alternative use first. This section will briefly sketch a model of typed feature structures using essentially just the tools outlined so far. The aim is to be able to use bits of channel theory that get at failed constraints to provide a theory of default feature structures. A default feature structure is one that contains local inconsistencies, and which a default reasoner can use to make reasonable assumptions in processing an ill-formed linguistic event in spite of the ill-formedness of the structure that corresponds to the event. Elsewhere, the same (but extended) machinery has been applied to the semantics of miscommunication in natural language dialog (Healey & Vogel, 1994a, 1994b).

Figure 5.3 depicts an inconsistent unification. The top two feature structures can be unified with the `noun` and `verb` nodes of the middle feature structure. This leads to the inconsistent feature structure in the bottom picture. Consider this as

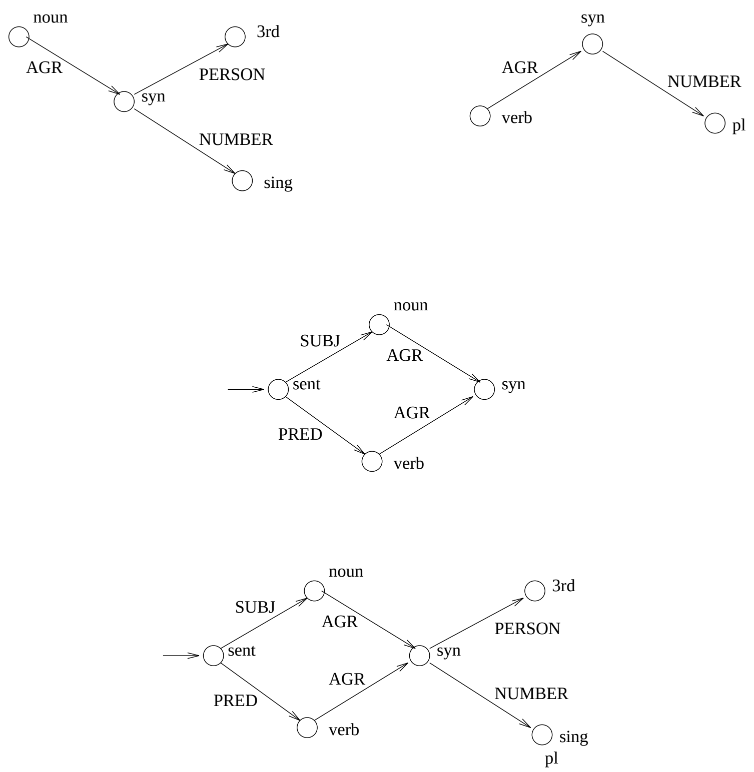


Figure 5.3: Unification

the structure behind the utterance, "The artist blow glass." Presumably there is access to the consistent information contained in the event, and when the event is embedded in a larger event for which the number of people in question is at stake, the other facts are also available to determine or decide by default whether one or more person is involved in the glass blowing, but it is known that some glass blowing is going on somewhere.

Classical Feature Structures

Carpenter (1992, p.36) presents classical feature structures (assume `Feat`, a set of features, and `Type`, a set of types):

> *A feature structure over* `Type` *and* `Feat` *is a tuple* $F = \langle Q, \overline{q}, \theta, \delta \rangle$ *where:*
>
> - Q*: a finite set of nodes rooted at* $\overline{q}$ *(see below)*
> - $\overline{q} \in Q$*: the root node*
> - $\theta : Q \rightarrow \texttt{Type}$*: a total node typing function*
> - $\delta : \texttt{Feat} \times Q \rightarrow Q$*: a partial feature value function.*
>
> *Let* $\mathcal{F}$ *denote the collection of feature structures.*
>
> The usual way to conceptualize a feature structure is as a labeled rooted directed graph where Q is the set of nodes, θ determines the labels on the nodes and where there is an arc from q to q' labeled by f, which we write as $q \xrightarrow{f} q'$ if $\delta(f, q) = q'$. In general we write:
>
> (5.8) $\quad q : \sigma \xrightarrow{f} q' : \sigma'$
>
> if there is an arc labeled with f from q to q', and the type assigned to q is σ and to q' is σ', or more formally, when $\delta(f, q) = q', \theta(q) = \sigma$ and $\theta(q') = \sigma'$.

The notation is suggestive of the channel theoretic interpretation. Features will be types of channels, instances of which connect nodes that are assigned types either directly or via the target classification that follows from the classification of a channel type with a particular constraint ($\overset{C}{\models}$).

Subsumption in feature structures is defined in terms of graph morphisms, and unification is information conjunction. In a classical feature structures, unification of two feature structures gives the least upperbound if one exists.

An example of a feature structure in graph notation, also taken from Carpenter (1992, p.37) is given in Figure 5.4

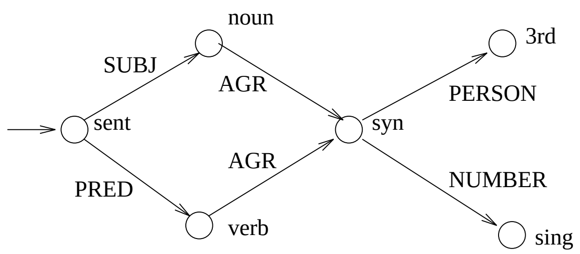


Figure 5.4: An Example Feature Structure

A Channel Theoretic Interpretation

In this section the same object is defined using channel theoretic notions from earlier in the chapter. Nodes will turn into sites, and constraints yield the types that will model structure types. This presentation is related to the one given by Vogel and Cooper (1994), however that work used a different formulation of channel theory. Channels model features. Let `Type` be a set of types, and let `Feat` be a set of channels.

Definition 48 *A feature structure of* `Type` *and* `Feat` *is a tuple* $F = \langle Q, \overline{q}, \theta, \Delta, \delta, \rangle$ *related to a classification domain* $\langle T, S, \models, \mapsto, \Longrightarrow, - \rangle$ *in the following way:*

- Q*: a finite set of sites rooted at* $\overline{q}$ *(see below)*
- $\overline{q} \in Q$*: the root token*
- *A set of types* $T = \mathtt{Type} \cup \Longrightarrow$
- *A set of tokens* $S = Q \cup \mathtt{Feat}$
- $\delta : \mathtt{Feat} \times Q \to Q$*: a partial function from feature value pairs into values defined such that* $\delta(f, q) = q'$ *iff* $q \overset{f}{\mapsto} q'$.
- $\Delta : Q \to 2^{\mathtt{Type}}$ *is a total classification function into sets of types, such that*
  1. $\Delta(\overline{q}) = \{\phi\}$ *iff* $\overline{q} \models \phi$
  2. $\forall q' \neq \overline{q}, \Delta(q') = \{\psi \mid \exists q, f : q \overset{f}{\mapsto} q',\ f \models \phi \Longrightarrow \psi$ *and* $\Delta(q) = \phi\}$.
- $\theta : Q \to \mathtt{Type}$ *is a partial classification function into types is defined such that* $\forall q, \theta(q) = \bigwedge(\Delta(q))$ *if* $\bigwedge(\Delta(q))$ *is consistent, and is undefined otherwise.*

Let $\mathcal{F}'$ *denote the collection of feature structures.*

So, here, a feature structure is conceptualized as a set of connections among sites. In particular, sites are typed via the constraints that classify the features which connect the sites. The relationship between features and constraints corresponds roughly to the appropriateness relation between types and features in Carpenter's presentation: $F \models T \Rightarrow T'$ iff features like F are appropriate for feature structures of type T and can have values of type T'. Note that $\Delta(q) = T$ if and only if there exists a feature (channel) $f \models F$ that is connected to q in a signaling relation such that $q \overset{F}{\models} T$. The θ mapping is a partial function related to Δ but which never classifies nodes inconsistently.

The equivalent to Carpenter's (5.8) is:

$$q \models \sigma \overset{f}{\mapsto} q' \models \sigma' \tag{5.9}$$

which means that there is a channel f which connects q to q', and the type assigned to q is σ and to q' is σ'. However it is the typing of nodes that follows from the indicating relations that is important here:

$$\begin{array}{ccc} \sigma & \Longrightarrow & \sigma' \\ C\|\!\!| & & C\|\!\!| \\ q & \overset{f}{\mapsto} & q' \end{array} \tag{5.10}$$

As before, proper subsumption will be a morphism, and unification will be information conjunction.

Definition 49 *$\sigma - \sigma'$ iff σ and σ' are incompatible.*

Figures 5.5 and 5.6 depict classification domains that model feature structures for the `NUMBER` feature. In the first case, the value of `NUMBER` is of type `Sing` and in the second case it is of type `Pl`. The dotted lines denote instances of the positive of-type relation ($\overset{\texttt{NUM}}{\models}$ and $\overset{\texttt{NOUN}}{\models}$). The typing of the feature itself is not actually denoted, although the indicating relation that types the feature is depicted.

Before defining unification, it is useful to illustrate how it works for feature-value pairs. The unification of feature-value pairs $q \overset{f}{\mapsto} q'$ and $r \overset{g}{\mapsto} r'$ will result in a new classification which identifies the tokens $(q = r, q' = r', f = g)$. The indicating relation in the result is the union of the indicating relations in each of the originals.

Note that Aït-Kaci's (1984) formulation admitted inconsistencies at nodes but then ruled them out via smashing to obtain bottom for the whole feature structure. In this model it is useful to leave the inconsistencies in. It's possible to

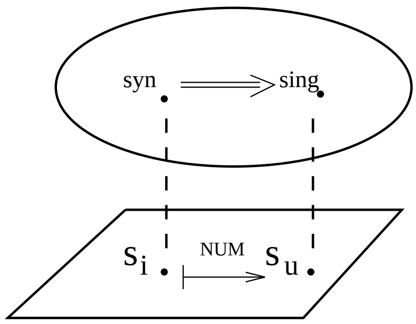


Figure 5.5: The `Number` Channel: Singular

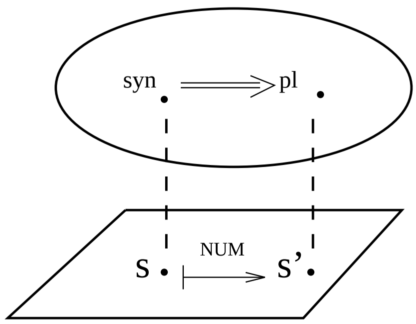


Figure 5.6: The `Number` Channel: Plural

look at just the restricted universe of consistent feature structures, but these are properly contained in a hierarchy of inconsistent structures. However, it is necessary to define a relative inconsistency ordering so that it is possible to refer to the minimally inconsistent structures.

Here it is necessary to introduce a channel theoretic concept which was not presented earlier; below are definitions for alternative sorts of parallel composition of channels $\|$.

Definition 50 *Parallel Composition*

- *A channel c is the* parallel composition *of c_1 and c_2 ($c_1 \| c_2$) iff for all sites $s_1, s_2 \in Q$ $s_1 \overset{c}{\mapsto} s_2$ iff $s_1 \overset{c_1}{\mapsto} s_2$ and $s_1 \overset{c_2}{\mapsto} s_2$.*
- *A channel c is the* consistent parallel composition *of c_1 and c_2 ($c_1 \| c_2$) iff for all sites $s_1, s_2 \in Q$ $s_1 \overset{c}{\mapsto} s_2$ iff $s_1 \overset{c_1}{\mapsto} s_2$, $s_1 \overset{c_2}{\mapsto} s_2$, and $c_1 \models T \Rightarrow T'$ and $c_2 \models T \Rightarrow T'$.*

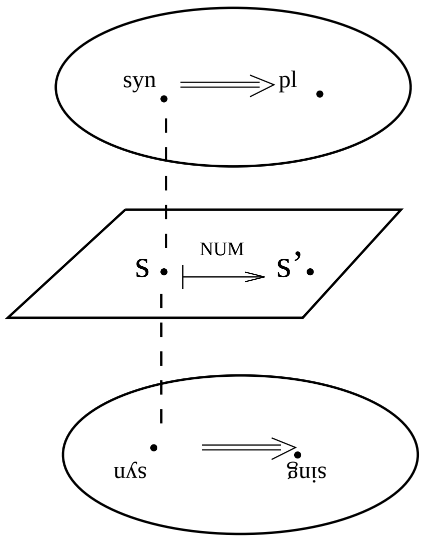


Figure 5.7: Restricted Classification of the Unification of Two Feature Structures

To model classical feature structures, it is required that the unification of two feature structures satisfy consistent parallel composition. But general feature structures need only satisfy unrestricted parallel composition. The resulting structures are potentially profoundly inconsistent. That is, each token can legitimately support a host of types, some of them conflicting. Figure 5.7 depicts a classification for the inconsistent feature structure resulting from the unification of feature structures represented in Figure 5.5 and Figure 5.6. Note that s' is shown to support both of the conflicting types `Pl` and `Sing`.

5.2.5 Basic and Composite Channels

The example application of modeling feature structures in channel theory leads to the need for composition of channels. The semantics of default inheritance reasoning, rather than requiring variations of parallel composition, will make crucial use of various serial composition operations. Barwise also provides a definition of serial composition of channels; it is useful to see how this compares with the definition required for the interpretation of inheritance:

Definition 51 *A channel c is the* putative serial composition *of n channels $c_1 \ldots c_n$ ($c_1; \ldots ; c_i; \ldots ; c_n$) iff for all sites $s_0, s_n \in S$, $s_0 \overset{c}{\mapsto} s_n$ if there are $n-1$ intermediate* *sites such that* $s_0 \overset{c_1}{\mapsto} s_1, \ldots s_i \overset{c_{i+1}}{\mapsto} s_{i+1} \ldots s_{n-1} \overset{c_n}{\mapsto} s_n$ $(1 \leq i < n)$.

Barwise actually refers to Definition 51 as serial composition simpliciter (and defines it instead as a binary operator), but given that this thesis requires a number of different potential forms of composition, it is useful to distinguish putative composites from those actually to be included in a particular model. Definition 51 generalizes to channel types in Definition 52.

Definition 52 *A channel type* C *is the* putative serial type-composition *of* n *channel types* $C_1 \ldots C_n$ $(C_1; \ldots; C_n)$ *iff for all* i, $1 < i \leq n$ *and for each* c^i *such that* $c^i \models C^i$ *there exists a putative serial composite* $c' = (c^1; \ldots; c^i; \ldots; c^n)$.

Definition 53 *Let* $c = (c_1; \ldots; c_i; \ldots; c_n)$ *be the putative serial composition of* $n \geq 2$ *channels such that*
$c_1 \models T^1 \Rightarrow T^i; \ldots; c^i \models T^i \Rightarrow T^{i+1}; \ldots; c_n \models T^{n-1} \Rightarrow T^n$, *then* $c \models T^1 \Rightarrow T^n$.

In what follows, various versions of serial composition will be more restricted in applicability than putative serial composition and will each have analogous generalizations to composition of channel types.

Given a composite channel $c = (c_1; ...; c_n)$, it is handy to be able to refer to the components of the channel $\{c_1, ..., c_n\}$. It is also useful to define the *front* of a channel as the sequence $\langle c_1, ..., c_{n-1}\rangle$ and the *back* of a channel as the sequence $\langle c_2, ..., c_n\rangle$ of the n channels that comprise a composite. Given a sequence of channels, $\langle c_i; ...; c_j\rangle$, it is meaningful to talk of channels composed *from* that sequence $\langle c_i; ...; c_j\rangle$ (i.e., $c = \langle c_k; ...; c_i; ...; c_j; ...; c_n\rangle$, $k \leq i \leq j \leq n$), channels composed from *behind* $\langle c_i; ...; c_j\rangle$ (i.e., $c = \langle c_i; ...; c_j; ...; c_n\rangle$, $i \leq j < n$), and channels composed from the *front* of $\langle c_i; ...; c_j\rangle$ (i.e., $c = \langle c_h; ...; c_i; ...; c_j\rangle$, $h < i \leq j$).

Consider the structure: $\langle S, T, \mathcal{C}, I, \models, \mapsto\rangle$. Let $\langle S, T', \mathcal{C}', I', \models', \mapsto'\rangle$ be the closure of the former structure under putative serial composition. It is useful to define some simple concepts from these two structures.

Definition 54 *Let the channels in* $\mathcal{C}$ *be* atomic *or* prime *channels. The composites in* $\mathcal{C}' - \mathcal{C}$ *are* nonatomic.

Definition 55 *Let c be a channel; PRIMES(c) denotes the set of atomic channels that comprise c.*

Definition 56 *Let c and c' be channels; c and c' are* site-equivalent *iff* $\forall s_1, s_2$, $s_1 \overset{c}{\mapsto} s_2$ *if and only if* $s_1 \overset{c'}{\mapsto} s_2$.

Definition 57 *An atomic channel $c \models T \Rightarrow T'$ is* basic *iff for any c' which is a nonatomic site-equivalent to c, c' and c support mutually inconsistent constraints.*

Consider (5.11) with $c^1, c^2 \models C$; $c^3, c^4 \models C'$; $c^5, c^6 \models C''$.

$$(5.11)\quad \boxed{\begin{array}{ccccccc}
T & \Longrightarrow & T' & & T' & \Longrightarrow & \neg T'' \\
\perp\!\!\!\perp\!\!\!\perp & & \perp\!\!\!\perp\!\!\!\perp & & \perp\!\!\!\perp\!\!\!\perp & & \\
s^1 & \overset{c^1}{\mapsto} & s' & & s' & \overset{c^3}{\mapsto} & s^2 \quad \models T'' \\
 & & & & & & \\
T & \Longrightarrow & T' & & T' & \Longrightarrow & \neg T'' \\
\perp\!\!\!\perp\!\!\!\perp & & & & \perp\!\!\!\perp\!\!\!\perp & & \perp\!\!\!\perp\!\!\!\perp \\
s^3 & \overset{c^2}{\mapsto} & s'' & \models \neg T' & s'' & \overset{c^4}{\mapsto} & s^4 \\
 & & & & & & \\
T & \Longrightarrow & \neg T'' & & T & \Longrightarrow & \neg T'' \\
\perp\!\!\!\perp\!\!\!\perp & & & & \perp\!\!\!\perp\!\!\!\perp & & \perp\!\!\!\perp\!\!\!\perp \\
s^3 & \overset{c^5}{\mapsto} & s^4 & & s^5 & \overset{c^6}{\mapsto} & s^6 \\
 & & & & & & \\
T & \Longrightarrow & T'' & & & & \\
\perp\!\!\!\perp\!\!\!\perp & & \perp\!\!\!\perp\!\!\!\perp & & & & \\
s^7 & \overset{c^7}{\mapsto} & s^8 & & & &
\end{array}}$$

Let $c' = (c^1; c^3)$, $c'' = (c^2; c^4)$; then, $c', c'' \models T \Rightarrow \neg T''$ (the same as c^5 and c^6). Also, $C^{\triangle} = (C; C')$. The atomic channels are $\{c^1, c^2, c^3, c^4, c^5, c^6\}$, and c', c'' are nonatomic. $\mathit{PRIMES}(c') = \{c^1, c^3\}$; $\mathit{PRIMES}(c'') = \{c^2, c^4\}$. The nonatomic channel c'' is site-equivalent to the atomic channel c^5 but not to c^6. Therefore, c^6 is basic but c^5 is not (and $c^1 \ldots c^4$ are all basic). Note that C'' and $C^{\triangle}$ both have a dual signal in s^7 through C'''.

Also consider a notion that can be defined in channel theoretic terms as *originality*. A channel is original if it is basic, or if it is not basic it is original if it supports the flow of information that is not supported by its constituent channels.

Definition 58 *An atomic channel $c \models T \Rightarrow T'$ is* original *iff*

1. *c is basic, or*

2. *c has a nonatomic equivalent decomposable into $\{c_1 \dots c_n\}$ such that for some $i, 1 \leq i \leq n$, $c^i \models C$ and C has a dual signal.*

In (5.11), although c^5 is not basic it has a nonatomic equivalent c'' which has a dual signal through c^7 in C'''. Thus, c^5 is original. The channels $c^1 \dots c^4, c^6$ are all original because they are basic. The channel c^7 is basic as well.

Definition 59 *An atomic channel type C is* original *if all c such that $c \models C$ are original.*

Definition 60 *A channel $c \models T \Rightarrow T'$ is* effectively original *iff*

1. *c is original, or*
2. *c is decomposable into atomic channels $c_1 \dots c_n$ such that each c_i is original.*

Definition 61 *A channel type C is* effectively original *iff each c such that $c \models C$ is effectively original.*

Effective originality extends the notion of original contributions of information from primes to putative composites. In the definition of composition offered below, composition with effectively unoriginal channels is disallowed. This is sensible, because, by definition, the same information is already conveyed elsewhere if a putative composite is effectively unoriginal.

Now it is possible define the directness of a channel (denoted, $|c|$) as the number of original channels in its putative composition; if c is a basic channel then $|c| = 1$. This admits the possibility of organizing channels into a specificity hierarchy. The one most useful for interpreting permission in H90 follow:

Definition 62 *Given two channels c_1 and c_2, each of the type $T_1 \Rightarrow T_2$ or $T_1 \Rightarrow \neg T_2$, c_1 is at least as direct as c_2 ($c_1 \preceq c_2$) iff $|c_1| \leq |c_2|$.*

Thus, one channel is more direct than another if it connects the same points and is composed from fewer original channels.

5.3 A Model of Inheritance Reasoning

Let ? be an inheritance network. Let M be a tuple: $\langle S, T, \mathcal{C}, I, \models, \mapsto \rangle$; S is a nonempty set of sites and T is a set of types. I is an interpretation function which assigns a unique type to each node in the inheritance network. For each link $n_1 \longrightarrow n_2$ in ? there is an informative channel type $C \in \mathcal{C}$ of the type $I(n_1) \Rightarrow I(n_2)$, and for each link $n_1 \not\longrightarrow n_2$ there is exactly one informative channel type C that supports the constraint type $I(n_1) \Rightarrow \neg I(n_2)$. Note that there are no constraints of the form: $\neg\ I(n) \Rightarrow \tau$ for any node n or any type τ. For each channel type $C, C' \in \mathcal{C}$, if C supports a constraint whose consequent type is identical to the antecedent of a constraint supported by C', then for some $c \models C$ and $c' \models C'$ there exists a site s which is a target to C and a source to C'. There are no other channels in M. A model of inheritance reasoning over ? is given by M', a tuple $\langle S', T', \mathcal{C}', I', \models', \mapsto' \rangle$ derived from M via the closure under the appropriate version of serial composition of channels (to be defined in the course of this section), with the assumption that no target to a channel is also a target to a channel composed from it. A path π through ? is supported by a particular reasoner if and only if $I(\,\mathit{firstnode}(\pi)) \Rightarrow I(\,\mathit{lastnode}(\pi))$ is a constraint supported by an informative channel type $C \in \mathcal{C}$ in the corresponding closure.

Proposition 8 *Given an inheritance network ?, π is a positive path in ? if and only if there exists an informative channel C such that*
$C \models\ \mathrm{I}(\,\mathrm{firstnode}(\pi)) \Rightarrow\ \mathrm{I}(\,\mathrm{lastnode}(\pi))$ *in M', and π is a negative path if and only if there exists an informative channel C such that*
$C \models\ \mathrm{I}(\,\mathrm{firstnode}(\pi)) \Rightarrow \neg\ \mathrm{I}(\,\mathrm{lastnode}(\pi))$ *in M'.*

Proof: (For π, a positive path; the proof for negative paths is symmetric.)

$\supset$ Suppose π is a path in ?; then each link λ in π has an informative atomic channel type C in M supporting the constraint: $I(\,\mathit{firstnode}(\lambda)) \Rightarrow I(\,\mathit{lastnode}(\lambda))$. Thus, if π is a direct link then $C \models\ I(\,\mathit{firstnode}(\pi)) \Rightarrow I(\,\mathit{lastnode}(\pi))$ is in M' trivially. If π is a compound path, then by the construction of M there is an intermediate site to each channel type such that the putative serial composition can be formed. Thus, if π is a positive path then $C \models\ I(\,\mathit{firstnode}(\pi)) \Rightarrow I(\,\mathit{lastnode}(\pi))$ is in M', and C is informative.

$\subset$ Suppose C is an informative channel type that supports the constraint $I(\,\mathit{firstnode}(\pi)) \Rightarrow I(\,\mathit{lastnode}(\pi))$ in M'. Then, by construction of

M, if C is atomic then π is a direct link in ?, and is thus a path in ?. If C is not atomic, then it has a putative serial decomposition into atomic channel types corresponding to the links in the chain forming π. Since there is no constraint of the form: $\neg\ I(n) \Rightarrow \tau$ for any node n or any type τ, π must be a well-formed path in ?. $\square$

5.3.1 Redundancy

Proposition 9 *Given an inheritance network ? without directly conflicting links and an inheritance model $M' = \langle S', T', \mathcal{C}', I', \models', \mapsto' \rangle$ closed under putative serial composition, a link $\lambda \in$? is topologically redundant iff* $\mathrm{I}($ firstnode$(\lambda)) \Rightarrow \mathrm{I}($ lastnode$(\lambda))$ *is a constraint supported by an informative channel type $C \in \mathcal{C}$, and C is not basic.*

Proof:

$\supset$ If λ is redundant with respect to a longer path π then λ and π have the same polarity and:

1. $\mathit{firstnode}(\lambda) = \mathit{firstnode}(\pi)$
2. $\mathit{lastnode}(\lambda) = \mathit{lastnode}(\pi)$

By construction of M', each link in π is interpreted by an informative constraint which is supported by a channel type in $\mathcal{C}$, and $I(\mathit{firstnode}(\lambda)) \Rightarrow I(\mathit{lastnode}(\lambda))$ is supported by the channel type C^λ. Also by construction of M', for each pair C^i, C^{i+1} $1 \leq i < n$, where C^i supports the constraint $I(\mathit{firstnode}(\lambda^i)) \Rightarrow I(\mathit{lastnode}(\lambda^i))$ for i, the i-th link in π, there exists $c^i, c^{i+1}, s, s^i, s^{i+1}$ such that $c^i \models C^i$, $c^{i+1} \models C^{i+1}$ and $s^i \overset{c^i}{\mapsto} s$ and $s \overset{c^{i+1}}{\mapsto} s^{i+1}$. Therefore, a putative serial composite C^π exists supporting the constraint:

$I(\mathit{firstnode}(\pi)) \Rightarrow I(\mathit{lastnode}(\pi))$, but because of the constraints (1 and 2) above,

$C^\pi \models I(\mathit{firstnode}(\lambda)) \Rightarrow I(\mathit{lastnode}(\lambda))$. Therefore, $c^\lambda \models C^\lambda$ has a nonatomic equivalent in $c^\pi \models C^\pi$; hence, c^λ is not basic.

$\subset$ If $I(\mathit{firstnode}(\lambda)) \Rightarrow I(\mathit{lastnode}(\lambda))$ is a constraint supported by an informative atomic channel type $C \in \mathcal{C}$, and C is not basic, then there exists $c' \models C'$ which is a nonatomic site equivalent to $c \models C$

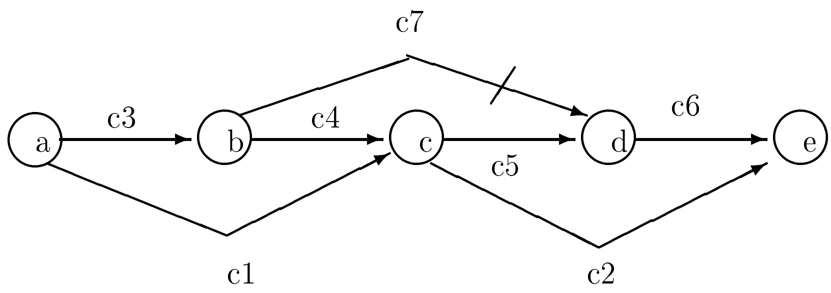


Figure 5.8: A Network with 7 Atomic Links, 5 Basic Links, 6 Original Links.

and the constraints supported by $c\&c'$ are mutually compatible. The construction of M' was based on the smallest set of channel types satisfying the assumptions (about channels of those types, signaling relations they participate in, and constraints they support), and there are no nonatomic channels in M. Thus, if there is a nonatomic channel C' in M' it exists because it is the putative serial composition of channels from M which were derived directly from ?, so there must be a path π such that $C' \models I(\mathit{firstnode}(\pi)) \Rightarrow I(\mathit{lastnode}(\pi))$. Similarly, if the atomic channel C in M exists, it is from the direct interpretation of a link λ' in ?. Then, if $c \models C$ and $c \models C'$ are site-equivalent it is because the endpoints of λ' and π coincide. From the interpretation function, and the fact that the constraints supported by the two channels are mutually compatible, this actually means that the shared sites are classified by identical types. If λ' is a direct link whose endpoints coincide with those of π then trivially the nodes of λ' are a subset of those of π. Therefore, λ' must be topologically redundant with respect to π.

□

It is easiest to understand the definitions of basic channels and of original channels given above, in terms of the network topologies that they interpret. Consider the graph from Figure 5.8. The network in this figure is labeled with channel types that interpret those links. The links ac and ce are interpreted by nonbasic channel types, and the rest are basic. The link b/d is interpreted as original since the node c offers a dual signal to c_2, and the link ce is original since b yields a dual signal for c_5, but the link ac is not original.

5.3.2 Interpreting Permitted Paths

Just as there is assumed to be a channel theoretic constraint underlying each link in an inheritance network, there is also assumed to be a composite constraint underlying each of the permitted paths in a network. Since permitted paths are defined by the exact specifications of the reasoning algorithm, this means that the channel theoretic interpretation of an inheritance network also provides an interpretation to the exact reasoning mechanism applied to that network. For instance, the definition of channel composition that is given in Definition 63 guarantees that composite channels exist for only those paths that are not preempted (by the H90 reasoner).

Restricted Skepticism, Off-Path Preemption

Finally, a definition of serial composition takes these notions into account. The serial composition of two channels can be formed as a channel if an intermediate site exists that supports the right types and if no more direct channel spans the two composed channels forming an antisignal to their composition; since this definition is particular to a model of H90 it is called *H90 serial composition*.

Definition 63 *A channel* $c \models T \Rightarrow T'$ *is the* H90 serial composition *of* $c_1 \models T \Rightarrow T''$ *and* $c_2 \models T'' \Rightarrow T'$ $(c_1;c_2)$ *iff*

1. *c is the putative composition of c_1 and c_2 where c_1 is an original channel or an effectively original H90 serial composite and c_2 is original, and*
2. *for any antisignal s through a putative channel c_i there exists a channel c_j through which s is a antisignal and a channel $c_\dagger$ supporting the same constraint as c, such that the following conditions are satisfied:*
   (a) $PRIMES(c_i) \cap PRIMES(c_j) \neq \emptyset$
   (b) $PRIMES(c_j) \cap PRIMES(c_\dagger) \neq \emptyset$
   (c) $c_\dagger \prec c_j$

The idea behind this restricted serial composition is that it is the complete interpretation of off-path preemption in H90. Essentially, there is no composite channel underneath paths composed from redundant links. Redundant links are unoriginal channels. Note that the $c\dagger$ mentioned in this definition could in fact be the putative composite c itself. If this occurs in the model of an inheritance network, then it is an instance of on-path preemption. Allowing $c\dagger$ to be distinct from c in its composition provides an interpretation to off-path reasoning. Note that the condition on non-empty intersection between the putative composite and antisignal channel (PRIMES$(c_j) \cap$ PRIMES$(c_\dagger) \neq \emptyset$) can be equivalently expressed in terms of set differences[5] in the following way:

(5.12) $\quad \text{PRIMES}(c_j)\backslash\text{PRIMES}(c_\dagger) \neq \text{PRIMES}(c_j)$

The equation in (5.12) indicates just that the prime channels in an antisignaling channel that are not in a channel of the original type cannot be the entire set of primes in the antisignaling channel: there must be an intersection. Expressing the intersection constraint in this way will be useful for later definition of on-path preemption.

For a given network ? , assume the existence of M and M' as constructed before. Let M^{H90} be the closure of M under H90 serial composition; M^{H90} constitutes a model of reasoning in H90 over ? . That is, if π is a positive path in ? permitted by H90, then $c : I(\textit{firstnode}(\pi)) \Rightarrow I(\textit{lastnode}(\pi))$ is an informative channel in M^{H90}, and if π is a negative path in ? permitted by H90, then $c : I(\textit{firstnode}(\pi)) \Rightarrow \neg I(\textit{lastnode}(\pi))$ is an informative channel in M^{H90}. Moreover, if $c : \phi \Rightarrow \psi$ is an informative channel in M^{H90} then a path π in ? supports that conclusion under the reasoning definitions of H90.

Proposition 10 *Given a network ? , if a negative path π in ? is off-path preempted by a positive path π' then in M'(?) there exists channel types C, C' (where possibly $C = C'$) and C'' as well as a site s which is an antisignal to all three, where*

- $C \models$ I(firstnode$(\pi)) \Rightarrow$ I($\neg$ lastnode(π)),
- $C' \models$ I(firstnode$(\pi)) \Rightarrow$ I($\neg$ lastnode(π)) *and*
- $C'' \models$ I(firstnode$(\pi')) \Rightarrow$ I(lastnode(π')),

with $c \models C$, $c' \models C'$ and $c'' \models C''$, and where the following conditions are satisfied:

1. *PRIMES(c) $\cap$ PRIMES(c') $\neq \emptyset$*
2. *$c'' \prec c'$*

[5] Recall the definition of the set difference operator $\backslash$, $A \setminus B = \{x : x \in A, x \notin B\}$.

Proof: Suppose the negative path π (σ/y) is preempted by the positive path π' $(\sigma' y)$ in ? using an off-path preempting reasoner. Then there exists a positive link $x \longrightarrow y$ and a positive path σ'' in ? such that

1. $firstnode(\sigma) = firstnode(\sigma'')$,
2. $lastnode(\sigma) = lastnode(\sigma'')$,
3. x occurs in σ'', and
4. $x \neq lastnode(\sigma'')$.

The construction of M' provides that there exists:

– $C \models\ I(\mathit{firstnode}(\pi)) \Rightarrow\ I(\neg\ \mathit{lastnode}(\pi))$,

– $C' \models\ I(\mathit{firstnode}(\pi)) \Rightarrow\ I(\neg\ \mathit{lastnode}(\pi))$ (where potentially, $C = C'$), and

– $C'' \models\ I(\mathit{firstnode}(\pi')) \Rightarrow\ I(\mathit{lastnode}(\pi'))$.

as well as $c'' \models C''$.

1. Suppose $\sigma \neq \sigma''$.
   Since σ and σ'' coincide at endpoints and are both positive, then both can be extended by the last link in π. Thus, there exist distinct negative paths: σ'/y and $\pi = \sigma/y$. Then by construction of M' there exist C and C' as described above and $C \neq C'$ which support the same constraint. Since σ'/y and σ/y have their last link in common, there exist $c \models C$ and $c' \models C'$ such that PRIMES$(c) \cap$ PRIMES$(c') \neq \emptyset$.
   Because π is preempted by $x \longrightarrow y$ and x occurs as a node in σ'' at some node that is not $\mathit{lastnode}(\sigma'')$ or $\mathit{firstnode}(\sigma'')$ (if it were $\mathit{firstnode}(\sigma'')$ then $\sigma = \sigma''$). Thus, σ'' must be a compound path and σ'' and π' coincide up to the node x. From x, π' has just the link $x \longrightarrow y$ and σ'' has one more positive link. Recall that σ'' can be extended by the negative link (the last link in π). Thus σ''/y has strictly more links than π'; therefore, in M', c'' has strictly fewer original links (if they were not original a different, more direct, channel would have been under consideration since a different path would have been at issue in the network) than c', so $c'' \prec c'$.
2. Suppose $\sigma \neq \sigma''$. Then $C = C'$.

The proof is much the same as above except that x can be the first node of σ'', as in the case of a preemptive link that spans the entire path it preempts. In that case, it is obvious that the corresponding channel types C'' and C' are such that C'' is more direct.

$\square$

The symmetric case of preemption of positive paths by negative paths also holds. Note that Proposition 10 does not include an only-if condition. That is because it is based on M' rather than M^{H90}. In M^{H90}, the only-if condition holds as well. Essentially, Proposition 10 means that if a path is preempted then in the interpretation closed under putative serial composition, the corresponding channel type is the less direct channel of conflicting channel types. Also note that the conditions on the conflicting channel types do not exhaust those of the 'conflicts that matter' clause of Definition 63. Consider a pair of conflicting channel types $C \models T \Rightarrow T'$ and $C' \models T \Rightarrow \neg T'$ in M'. If they satisfy the constraints in Proposition 10 with $c_j \models C$ and $c_\dagger \models C$, and additionally PRIMES$(c_j) \cap$ PRIMES$(c_\dagger)$ $\neq \emptyset$ then the corresponding conflicting paths π and π' in ? share only their endpoints. If C and C' are not in M but are in M' then they are putative serial composites, and thus π and π' are compound paths. This means that the neither path is permitted, and that clause of Definition 63 prevents either putative serial composite from being an H90 serial composite. Thus, neither channel type is in M^{H90}. H90 serial composition is a binary operator (just as is path construction in H90), which assumes that the first channel type is either original, or an H90 composite (thus, a channel type from M^{H90} and not just M'), and that the second channel type is original. This entails that if a channel type C is in M^{H90} (and not in M) it corresponds to a path in ? composed of a permitted path extended by a single link where the resulting path is neither preempted nor conflicts with any path. Therefore, Proposition 11 holds.

Proposition 11 *Given an inheritance network ?, π is a positive path in ? if and only if there exists an informative channel C such that*
$C \models \mathrm{I}(\text{firstnode}(\pi)) \Rightarrow \mathrm{I}(\text{lastnode}(\pi))$ *in* M^{H90}, *and π is a negative path if and only if there exists an informative channel C such that*
$C \models \mathrm{I}(\text{firstnode}(\pi)) \Rightarrow \neg\, \mathrm{I}(\text{lastnode}(\pi))$ *in* M^{H90}.

As in Chapter 4, modifications to the basic definition which provides an interpretation to the H90 system can be provided to create interpretations for the other possible modes of reasoning. Recall that the H90 system is restrictedly skeptical (hence skeptical and not credulous), utilizing off-path preemption, and exhibiting stability when faced with topologically redundant links.

Fully Skeptical Reasoning

Definition 63 proffered an interpretation for restricted skeptical reasoning. As described in Chapter 2, restricted skepticism is manifest in relation to nested ambiguities in an inheritance graph. Essentially, a fully skeptical reasoner averts conclusions when putative paths conflict; restricted skepticism requires that subpaths of putative paths themselves be permitted before accepting the longer paths as potential conflicts. It is easily seen that the interpretation given here addresses this by distinguishing putative composition from H90 serial composition. Thus, to achieve an interpretation for fully skeptical reasoning it is necessary to just relax the restrictions on H90 serial composition. The result is stated in Definition 64 in which the first condition on channel composition is made nonrecursive.

Definition 64 *A channel $c \models T \Rightarrow T'$ is the* H90 fully skeptical serial composition *of $c_1 \models T \Rightarrow T''$ and $c_2 \models T'' \Rightarrow T'$ $(c_1;c_2)$ iff*

1. *c is the putative composition of c_1 and c_2 where both channels are effectively original, and*
2. *for any antisignal s through a putative channel c_i there exists a channel c_j through which s is a antisignal and a channel $c_\dagger$ of the same type as c such that the following conditions are satisfied:*
   (a) *$PRIMES(c_i) \cap PRIMES(c_j) \neq \emptyset$*
   (b) *$PRIMES(c_j) \cap PRIMES(c_\dagger) \neq \emptyset$*
   (c) *$c_\dagger \prec c_j$*

Proposition 12 *Given an inheritance network ?, π is a positive path in ? if and only if there exists an informative channel C such that*
$C \models \mathrm{I}(\,\mathrm{firstnode}(\pi)) \Rightarrow \mathrm{I}(\,\mathrm{lastnode}(\pi))$ *in* $M^{H90FullSkeptic}$, *and π is a negative path if and only if there exists an informative channel C such that*
$C \models \mathrm{I}(\,\mathrm{firstnode}(\pi)) \Rightarrow \neg\, \mathrm{I}(\,\mathrm{lastnode}(\pi))$ *in* $M^{H90FullSkeptic}$.

On-Path Preemption

Both Definition 63 and 64 interpret off-path preemption in inheritance reasoning. Off-path preemption allows one path to be preempted by the fact that another path with the same endpoints and at least one link in common is also preempted by a "more specific" path of a conflicting type.[6] On-path preemption defines specificity in such a way that paths that have a diverging set of nodes cannot preempt each other.

Definition 65 *A channel* $c \models T \Rightarrow T'$ *is the* H90 on-path preemptive serial composition *of* $c_1 \models T \Rightarrow T''$ *and* $c_2 \models T'' \Rightarrow T'$ $(c_1;c_2)$ *iff*

1. *c is the putative composition of c_1 and c_2 where both channels are original or effectively original H90 serial composites, and*
2. *for any putative channel c_i through which there exists antisignal s to the putative channel c, the following conditions are satisfied:*
   (a) $\mid PRIMES(c_i) \setminus PRIMES(c) \mid = 1$
   (b) $c \prec c_i$

Definition 65 interprets on-path preemption since it does not consider the composition that creates the channel c to depend on other channels $c_\dagger$ of the same type as c; thus, the existence of an interpretation for a path depends crucially on a channel underneath it that itself preempts any channels that convey antisignals. The essential fact of preemption is that it involves one path being more specific than another conflicting path—this is captured in the directness relation—and when the preemption is on-path it is entailed that the conflicting path can differ by at most one link, or else there is simply an ambiguity and not a preemption.

Proposition 13 *Given an inheritance network ?, π is a positive path in ? if and only if there exists an informative channel C such that*
$C \models \mathrm{I}(\mathrm{firstnode}(\pi)) \Rightarrow \mathrm{I}(\mathrm{lastnode}(\pi))$ *in* $M^{H90OnPath}$*, and π is a negative path if and only if there exists an informative channel C such that*
$C \models \mathrm{I}(\mathrm{firstnode}(\pi)) \Rightarrow \neg\, \mathrm{I}(\mathrm{lastnode}(\pi))$ *in* $M^{H90OnPath}$.

[6]The scare quotes are important because this is actually offering a structural definition of what more specific means; it is not assuming that the conflicting path is more specific for other reasons.

Combinations

Definitions 65 and 64 each change just one parameter to interpret a version of H90 with the corresponding reasoning method. However, it is of course possible to combing fully skeptical and on-path reasoning, as in Definition 66.

Definition 66 *A channel* $c \models T \Rightarrow T'$ *is the* fully skeptical, on-path preemptive serial composition *of* $c_1 \models T \Rightarrow T''$ *and* $c_2 \models T'' \Rightarrow T'$ $(c_1;c_2)$ *iff*

1. *c is the putative composition of c_1 and c_2 where both channels are effectively original, and*
2. *for any putative channel c_i through which there exists antisignal s to the putative channel c, the following conditions are satisfied:*
   (a) $\mid PRIMES(c_i) \setminus PRIMES(c) \mid = 1$
   (b) $c \prec c_i$

Instability

It is difficult to find a succinct label for reasoners that embody instability since they have been frowned upon since Touretzky introduced the inferential distance ordering to "correct" the shortcomings of shortest path reasoning when "redundant" links are added to a network. Using a shortest path reasoner, a network in which a definite conclusion can be reached, when augmented with links derivable from subpaths, ceases to yield definite conclusions. The inferential distance ordering provides a way to identify which links are redundant in the sense of providing no more information than which can be derived from subpaths, and ignores them. However, this loses sight of the difference between derived links and links that do genuinely represent original information not contained in longer paths. In this section we provide an interpretation for H90 style reasoners which do not disregard the contributions of explicit links. Essentially, the difference is a relaxation of the requirement that the component channels in a composition be original or effectively original. This form of composition is labeled *literalist serial composition* to capture the intuition (in networks) that explicit links can be taken literally as providing information not contained in derived links and (in channels) that prime channels convey information that is not wholly captured by composites of the same type.

Definition 67 *A channel* $c \models T \Rightarrow T'$ *is the* literalist serial composition *of* $c_1 \models T \Rightarrow T''$ *and* $c_2 \models T'' \Rightarrow T'$ *(*$c_1;c_2$*) iff*

1. *c is the putative composition of c_1 and c_2, and*
2. *for any antisignal s through a putative channel c_i there exists a channel c_j through which s is a antisignal and a channel $c_\dagger$ of the same type as c such that the following conditions are satisfied:*
   (a) *PRIMES(*c_i*)* $\cap$ *PRIMES(*c_j*)* $\neq \emptyset$
   (b) *PRIMES(*c_j*)* $\cap$ *PRIMES(*$c_\dagger$*)* $\neq \emptyset$
   (c) $c_\dagger \prec c_j$

Credulity

Finally, note that all of the preceding definitions have given channel theoretic interpretations to composition underneath skeptical reasoners. Skeptical reasoners rule out definite conclusions when faced with unadjudicatablely conflicting paths. The corresponding channel theoretic interpretations have withheld the existence of composite channels that would have composite antichannels. This section provides an interpretation for credulous reasoning. In credulous reasoning it is assumed that in some extensions one of the conclusions is applicable and in other extensions, the other.

Definition 68 *A channel* $c \models T \Rightarrow T'$ *is the* credulous serial composition *of* $c_1 \models T \Rightarrow T''$ *and* $c_2 \models T'' \Rightarrow T'$ *(*$c_1;c_2$*) iff*

- *c is the putative composition of c_1 and c_2, where both channels are original or effectively original serial composites.*

Note that Definition 68 is virtually equivalent to the the definition of putative serial composition alone (see Definition 51) except for the restriction of effective originality which admits an interpretation for preemption into even credulous reasoning. An interpretation of full credulity obtains by releasing that restriction as well.

5.4 Conclusions

Channel theory provides a formal framework which admits distinctions of fine granularity attuned to the needs of providing semantics to the family of inheritance reasoners under consideration. The sanctioning of a path in a reasoner is interpreted by the existence of an informative constraint supported by either an atomic or composite channel in the corresponding closure of the atomic model. In using inheritance reasoning to ask if something represented by some node has a property represented by some other node the question is posed in terms of the existence of a permitted path between the two nodes. If there is a path and it is positive, the object represented by first node is assumed to have the property represented by the second node, and if the path is negative, then the object is considered to have the antiproperty. In the interpretation provided here, the existence of a path between the two nodes is interpreted as by the existence of a natural regularity of the denoted kind. That is distinct from the question of whether the object being considered does indeed behave in accordance with that natural regularity. It could still be an exceptional object of some sort. To answer questions about objects in the semantics, it is necessary to simply look at the types that classify it. In the case that certain types don't classify it explicitly then it is possible to perform a leftwards projection of the types that the object participates with in constraints (see §5.2.3). If leftwards projection simpliciter is used (in conjunction with H90 serial composition), the conclusions of skeptical inference are available. If, instead, rationalized leftwards projections are utilized then the model interprets the multiple extensions of credulous reasoning.

This chapter has provided an interpretation of certain kinds of inheritance reasoners in the family of path-based reasoners defined by Touretzky et al., using the tools provided by channel theory. In summary, the interpretation identifies the default links of an inheritance network with channels that admit information flow. Paths permitted by reasoners are identified with composite channels in the corresponding model. The restrictions on path permission imposed by particular inheritance proof theories are modeled by restrictions on the composition of channels. To my knowledge, this is the first time that a unified semantics for this particular family of reasoners has been presented. Past work has tended to either provide a semantics to a particular system (Touretzky, 1986), or to work from a semantic framework to a particular inheritance logic whose proof theory approximates reasoning in the desired models (Delgrande, 1988; Boutilier, 1989; Veltman, 1994). An advantage of the channel theoretic approach is that in providing a unified framework it facilitates argument about the intuitiveness of different proof theories for inheritance, by giving their respective assumptions about conclusions based on topological considerations clear counterparts in the semantic model. For example, it identifies the arguable counterintuitiveness of labeling some links of in a certain topological configuration redundant and others in the same topological configuration preemptors by using an axiom, which indeed stipulates that precedence works in the opposite way than it usually does when a single link is composed in parallel with a composite link of the same polarity between the same endpoints. This is the principle of redundancy.

Other work in giving a semantics to inheritance reasoning explains proof theoretic differences in terms of different "process models" without actually stating what the different process models are (cf. Dimopoulos, 1992). By stipulating the actual conditions under which channels may compose, this work can be seen as giving a "process model". However, the basic principle is that channels are objects that offer exactly the right level of description for modeling inheritance proof theory. And, although the model has its drawbacks, it does offer a successful first pass at modeling inheritance reasoning; it provides a starting point for further exploration of channel theoretic models. For instance, there does seem to be hope that this approach will permit modeling of cyclic networks as well as acyclic nets (the models constructed in this essay assumed, as did the proof theories addressed, that nets are acyclic), since channel theory has been used as a general theory of information flow to characterize what is happening in situation semantics (Barwise, 1993), and situation semantics has been given applications in other puzzles of circularity (Barwise & Etchemendy, 1987). This would be extremely useful since the proof theory of inheritance has not fully addressed circularity since no intuitive semantic interpretations have been available.

Clearly, however, this chapter represents only the start of a proper analysis of the semantics of inheritance reasoning. The family of reasoners under consideration needs to be broadened by incorporating other axes of proof theoretic variation. There is not room within the scope of this thesis to provide a semantics for all of the possible systems outlined in Chapter Four of this thesis, so extensions to this work should address that omission. Additionally, this work would benefit tremendously from detailed comparison of other frameworks for giving semantics to default reasoning, including minimal models accounts for inheritance networks (Boutilier, 1991, 1993), as well as semantics for other default systems like default logics (Przymusinski, 1989), update semantics (Veltman, 1991, 1994), and Kripke semantics for conditional logics (Delgrande, 1990) (see Chapter 2 for an outline of the approaches taken by Boutillier, Veltman and Delgrande). Given that inheritance networks offer a naive semantics for natural language generics, it would be quite useful to compare properties of the semantics provided here for inheritance reasoning to the model of natural language generics provided directly in channel theory by Cavedon and Glasbey (1994). Default inheritance links make typicality a primitive, thus are arguably an instance of the predicate operator *generally* or *usually* as considered, for instance, by (Schubert & Pelletier, 1987); however, inheritance reasoning is more concerned with chaining generics together into arguments than with isolating when it is appropriate to have an inheritance link as the interpretation of some sentence. Given the way the semantics of reason is modeled in channel theory here, this means that the semantics of the natural language statements themselves, as in Schubert and Pelletier's (1987) example *Dogs are intelligent* (p. 429) are not based on there being "'sufficiently many' realizations" that make it true, but on the existence of an information channel which enables the classification of some situation as if the statement were true.[7] As Cavedon and Glasbey (1994) point out as an advantage of their model of generics, this means that channel theoretic approaches can thus provide a coherent interpretation of sentences like *John shoots burglars* which would be true if John carried a gun and intended to make use of it without ever the occasion actually arising. It would be enlightening to establish the exact relationship between their model of generics and the semantics of inheritance logics provided here.

[7]In fact, it would be quite interesting to explore at greater length relationship between the channel theoretic approach and Schubert and Pelletier's (1987), since theirs does make critical use of distinguishing extensions of predicates from sets of objects whose evaluation with respect to predicates yields TRUE. That is, they say that a kind is in the extension′ of some predicate iff 'sufficiently many' instances of the kind are in the extension of the predicate (where extension′ is the concept they are trying to define). Thus, there is a similar use of types at work, one which admits defeasibility.

Chapter 6

Discussion

6.1 Summary

This thesis has addressed the problem of default inheritance reasoning from three main angles: proof theory, psychological plausibility, and semantics. Chapters Two and Four addressed mainly proof theoretic aspects, but included complexity analysis and implementations of inheritance reasoners as well. Chapter Three offered an evaluation of the plausibility of inheritance reasoning as a particular model of human reasoning with natural language generics. Chapter Five presented a semantics for inheritance reasoning within a channel theoretic framework. The task of this short chapter is just to recapitulate and critique the main points of the thesis. The chapter also suggests some promising ways to develop this research further.

6.2 Contributions

6.2.1 Proof Theory, etc.

Chapter Two introduced the main concepts from path-based default inheritance reasoning that were important throughout the thesis. The main properties of interest are defeasibility (of course), transitive reasoning, negative reasoning, path preemption, redundancy, degrees of skepticism and chaining direction. The chapter also provides a summary of the three main attempts in the literature to provide a semantics to inheritance, all motivated from the direction of the framework used for the semantics (Veltman, 1994; Delgrande, 1990; Boutilier, 1989). The result is that both Veltman (1994) and Delgrande (1990) provide a semantics to only a single inheritance system each, and both are only approximate because neither provides the proof theory for the corresponding inheritance system that does reach all and only the conclusions entailed in the semantics. Boutilier (1989) does provide both semantics and the corresponding inheritance reasoner, but again, only for one particular system. To my knowledge, there has not been a general semantics in which the variations among inheritance can be identified. The main contribution of this chapter to the literature is a re-assessment of the complexity results which have shown downwards reasoning to be intractable (Selman & Levesque, 1989, 1993). In fact, these results pertain only to a double chaining variety of inheritance, not to a single chaining variety of downwards reasoning symmetric to the upwards chaining of H90. Chapter Four provides formal definitions for path-based inheritance reasoning including the axes of variation named above. Some additional systems are also presented, corresponding to alternative conceptions of the sort of conclusions that are reasonable to draw (some based on the data obtained in the experiments reported in Chapter Three), including the introduction of a version of downwards reasoning that is symmetric to H90's upwards chaining in putative path construction. To my knowledge, apart from previous collaborative work involving the author, this is the first time that the proof theoretic possibilities outlined by Touretzky et al. (1987) have been formally defined within the same framework, admitting more exact comparison among them. Only initial steps in the direction of actually making comparisons were taken, for instance, considering the relative number of paths preempted by the various systems. An advantage of presenting the reasoners within a uniform style of definition, in addition to making it easier to locate the parameters that can be tweaked to yield different reasoners, is that it lends itself easily to actually implementing those systems in a similarly parameterized fashion. Prolog implementations are presented and described in this chapter. Using Prolog as the implementation language has the advantage of making verification easier due to the close correspondence between the text of the code and the text of the definition (subject to considerations like subgoal ordering, etc.).

6.2.2 Psychological Plausibility

Chapter Three presents some informal experiments designed to test whether a defeasible interpretation of implication provides a better fit than classical logic does with human application of *modus tollens*. It has been seen that classical logic has little predictive power of people's response to those logic problems. If classical logic is a good theory of reasoning competence, then people diverge considerably in their performance. If classical logic offers a good theory of truth relations but is not applicable as a theory of rational competence, then other logics which provide a closer match to peformance should be considered. This thesis adopts the latter perspective, and Chapter Three proceeds to investigate specific features of inheritance reasoning as an alternative competence model and the degree to which people satisfy them. A wide range of subjects (some obtained using Internet access, a new and powerful mode of access to the general population (Laurent & Vogel, 1994)) were each presented forty problems comprised of inheritance graphs, either in graphic notation or just as the equivalent set of sentences. The defaults were expressed about abstract categories to prevent uncontrollable interaction of world knowledge. Mainly the results here are negative for the specific proof theoretic features considered elsewhere in the thesis. However, the prediction that people would make judgements that suggested they sanctioned conclusions corresponding to paths through a network (transitivity) was upheld. Preemption by a single link (not by paths) was also a rather good predictor, considering just the subjects that did not rate the relevant problems ambiguous. In fact, the majority of subjects rated complex problems indeterminate; for problems that might have discriminated, for instance, whether restricted or full skepticism is more appropriate, people overwhelmingly classified them as indeterminate. Additionally, it seemed that path counting, a weighing of the raw number of arguments for or against a conclusion, actually supplied a better predictor of subject responses than most extant proof theories. This is important, because it is direct evidence against the idea of identifying links as redundant based on topological configuration: if a 'redundant' link contributes to an argument that sways the path count and, hence, the decision, then it cannot be semantically redundant.

An interesting negative result is the suggestion that people reason with negative defaults in a way wholly unpredicted by the literature, admitting definite conclusions about paths containing nonfinal links. Such conclusions are certainly not statistically valid, and it is surprising that the result obtains. (6.1) gives an example with concrete concepts.

(6.1) $penguin \not\longrightarrow flier \longrightarrow winged$

The motivation for allowing a negative link only as the last link of a negative path is that if all that is known is that two classes are disjoint, then that gives one no information about the relationship between either of those two classes and classes that the other might be positively related to. This is illustrated with Venn diagrams in Figure 6.1: if all that is known is that As are not Bs and that Bs are Cs then that is not enough information to classify the relationship between As and Cs—the relationship could go either way, as the diagram indicates. Perhaps defea-

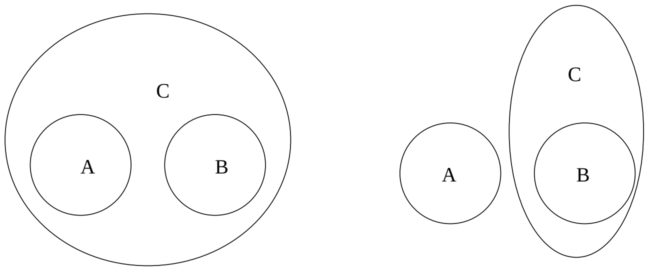


Figure 6.1: $A \not\longrightarrow B \longrightarrow C$

sibility creates a different situation from the crisp set-theoretic one. In any case, if people responded to concrete categories the way they did to abstract ones, they would be inclined to infer from *just* the information in (6.1) that penguins typically do not have wings. Thus, the sense is that negativity propagates down a path. It is not clear whether multiple negatives intensify or cancel each other. Note that one implication of accommodating in the proof theory nonfinal negative links in a path is that negative cycles (Geffner & Verma, 1989) become problematic again. It is interesting to consider the conduit metaphor for channels which interpret paths in the semantics: there is a natural metaphor for negative links being the last link in a path (negative information blocking the pipe) as well as for intensification (the negative features accumulating in the pipe), but less so for double-negative cancellation.[1] Of course, more work is required to ascertain whether that negative typicality might really be the denial of a positive one. However, subjects were provided with an option that did deny the positive assertion. Note that if other information besides what is represented in (6.1) is known, then there is an additional link which should have been represented in the network—presumably one which asserts that penguins typically have wings. Thus, any example like (6.1), which is designed to show that chaining past negative links is a bad idea, actually works as such an example on the basis of a positive link that the reader knows should rightfully be present. On the other hand, the complementary set of examples are those in which it works out just fine to reach the negative conclusion as in (6.2).

(6.2) $cat \not\rightarrow flier \longrightarrow winged$

While this point was not made in the chapter itself, it is illustrative of the way that the large body of data amassed in these experiments can be used to inform inheritance proof theory, based on what people find reasonable to conclude.

6.2.3 Semantics

The fifth chapter of this thesis introduces the channel-theoretic wherewithal later used to supply a semantics for a family of inheritance reasoners, a subset of those described in Chapter Two and Four. Channel theory is a new mathematical framework originating with work in situation theory (Barwise, 1989; Seligman, 1990; Barwise, 1991, 1993; Barwise & Seligman, 1994), as a way of expressing natural regularities. Regularities are modeled with real objects, channels, that connect situations about which information flows. The framework presented in Chapter Five develops a version of the theory presented by Barwise (1991). A brief example in that chapter shows how channels can be used as a start for a theory of nonclassical feature structures. The model developed within this channel theoretic framework takes advantage of the fact that generics and defaults are kinds of regularity and uses channels as a semantic object with which to model inheritance. Atomic channels are distinguished from composite channels, and inheritance graphs are given

[1] Of course, one should always be cautious about taking speculative metaphors as fact, but it is an interesting observation nonetheless.

atomic channel theoretic interpretations. Different inheritance reasoners license different paths, and correspondingly different restrictions on the composition of channels yield semantics for those reasoners. A path is supported by a reasoner just when there is a channel corresponding to that path in the model when closed under the appropriate composition operator. Chapter Five presents a semantics for H90, an upwards chaining, off-path preempting, restricted skeptical reasoner. It also shows the forms of composition required to obtain the sort of downwards chaining introduced in Chapter Four, as well as on-path preemption, full skepticism, and alternative conceptions of 'redundant' links. To my knowledge this is the first parameterized semantics for a family of inheritance reasoners. Dimopoulos (1992) has also considered the semantics of more than individual systems, but stipulated that 'process models' underlie each without specifying what those process models are. I argue that channels provide exactly the right level of description in order to specify a semantics for inheritance that makes the correspondences between proof theory and semantics explicit.

6.3 Further Work

In all three categories where this thesis has made positive contributions to the literature, there is scope for further development.

6.3.1 Proof Theory, etc.

Having spelled out a parameterized definition for one family of inheritance reasoners, it would be extremely useful to broaden the coverage and to add greater depth of analysis. Chapter Four demonstrated how reasoners could be related to each other in terms of properties like 'tendency to indeterminacy' which Chapter Three revealed to be a feature of human reasoning with defaults. There are almost certainly a great many other interesting properties to consider in this fashion, by identifying the minimal differences that achieve mutation of one reasoner into another. Possibly, more important properties are to be discovered, such as a deeper insight into the complexity issues and exactly what proof theoretic features take a reasoner from one complexity class into another. While Touretzky et al. (1987) have outlined a space of variations, and this thesis has expanded on that and within a formal setting, it has not taken full advantage of the formalization in relating the systems to each other. That is for future work, and this thesis has set the foundations for it.

6.3.2 Psychological Plausibility

A criticism that can be leveled against the investigation of the psychological plausibility of inheritance reasoning are about the actual administration of experiments. Internet experimentation did supply wide access to subjects, but coordinating materials attuned to the technical requirements of networked subjects (for instance, whether or not graphics could be printed or viewed or only text files) and timing constraints made it difficult to obtain a proper balance of subjects in each condition. However, this just entailed employing a statistical test which is insensitive to those imbalances under certain other conditions. Log linear analysis provided that flexibility, and tests of degree of association offered back-up checks on potentially spurious significance. Moreover, as pointed out in the introductory chapter to this thesis, there has not before Elio and Pelletier (1993) and Hewson and Vogel (1994) been an investigation of the psychological plausibility of inheritance reasoning, so any advances in this area are a contribution to the literature. Chapter Three, unfortunately, did not offer room to analyze the mass of data accumulated in the experiments to the fullest degree that is possible. Cluster analysis would unearth interesting information about styles of reasoning and consistency of individual members within a partition. Any number of other investigations of problems in the data set could also be conducted as a preliminary step to more refined experiments. It did turn out that forty questions was excessively demanding of subjects' energy, especially without remuneration. Future work in this area should use more constrained experimental instruments. However, an advantage of having used the extensive questionnaire and large sample sizes initially is that there is now a substantial body of data that can be explored to the greatest extent possible to extract information about other proof theoretic issues that are represented in the problem set. It is clear even from the example given in §6.2.2 that there is much to be gained by using empirical results to inform the proof theory of inheritance.

It is also clear that a great deal more research in this area is required to clarify some of the results. Now that a methodology exists based on comparing responses to minimal pairs of sets of defaults about abstract categories, this can be applied further using questionnaires designed to test specific issues more fully. For instance, the effect of path length was investigated only up to three links—even though longer paths were represented in the problem set, they were complicated for other reasons and could not reliably inform comment on path length. Another issue, which has been addressed in a subsequent research proposal based on this material, is the conjecture that the overwhelming tendency towards indeterminacy is a function of the task design. Recall that subjects were presented with sets of defaults and then asked to state about two of the classes (always the extreme endpoints X and Y in the graphic translation), whether, X are typically Y, typically not Y, both typically Y and typically not Y, neither typically Y nor typically not Y or whether they didn't know. For analysis purposes the last category was excluded and the two categories before that were grouped together as the general category 'indeterminate'. In particular, there were no consequences of having made a decision in a particular way, thus no pragmatic need to rate complex graphs as anything other than indeterminate. It would be quite useful in subsequent research to design a follow-on task so as to encourage more determinate answers. A follow-on task which infuses the original task with goal-directedness may reach different results, less indeterminacy. It would also be useful to explore negativity (as described earlier) and indeterminacy in greater detail to ascertain whether people really do entertain directly conflicting defaults and the extent to which negative classifications are mistaken for denials of positive classifications.

In fact, having acknowledged that the inheritance literature could benefit from giving greater attention to the way people actually reason with generics and defaults, it becomes obvious that there is quite a lot more to be investigated than can be enumerated in this final chapter. Chapter Three of this thesis, following on from the work of Hewson and Vogel (1994), has presented one methodology for exploring these issues (using abstract categories, pairwise comparisons of responses, etc.), that complements the methodology proposed by Elio and Pelletier (1993) (who use real-world problems, focus on the prediction that one exception to a default is irrelevant to the applicability of the default to a distinct object, etc.). Subsequent research in this area will benefit from having contrasting methods to work with.

6.3.3 Semantics

The semantics for inheritance reasoning provided in Chapter Five needs to be related in detail to other semantic approaches. To an extent, this is the task of channel theory since as a mathematical framework it is in part responsible for indicating its equivalences with other frameworks. But largely such equivalences are apparent through specific tasks, like the one at hand in providing a semantics to a family of inheritance reasoners. The difficulty in establishing equivalences follows from there being no other framework applied to the same task. While, for instance, possible worlds and minimal models have been used to give semantics to specific inheritance reasoners, the choice of inheritance reasoner was somewhat arbitrary and followed essentially from the decisions made in the modeling framework, and they have not been extended to give semantics to other inheritance reasoners. It remains possible that establishing an equivalence with respect to the inheritance reasoners that those systems give semantics to would, through the generality of the channel theoretic approach, suggest how to modularize those approaches as well. Thus, there is real value in constructing a detailed comparison between the modeling frameworks as applied to the task of giving a semantics to default inheritance reasoning.

Further, as with the proof theoretic and empirical investigations reported in this chapter, there is room to utilize the same framework to broaden the class of inheritance reasoners under consideration. Moreover, given the work of Cavedon (1995), there is an additional foothold (beyond the one through Delgrande's (1988, 1990) conditional logic and its application as a semantics for inheritance) on establishing connections between channel theoretic models and conditional logics, and thence to other formal systems that model defaults. It would be a rather illuminating exercise to tie the present work on giving a semantics to inheritance reasoning, since inheritance does offer a naive natural language semantics for generics, with work that gives natural language generics semantics directly within channel theory (Cavedon & Glasbey, 1994) and with the literature on the natural language semantics of generics. This offers the promise of a coherent and complete picture of the role of generics and reasoning and seems within reasonable reach.

6.4 Final Remarks

This thesis has been a tripartite investigation of path-based default inheritance reasoning. Clearly there is more to inheritance than just those systems considered here, but the systems chosen were not arbitrary. They are those which have been proposed within the literature, related reasoners, and reasoners suggested by the way people respond to sets of abstract defaults. Moreover, their proof theory has been interrelated through a parameterized set of definitions (implementations for some of them supplied) as well as with a semantics that provides models for the family. The thesis urges strongly that further energy be devoted to empirical study of the way people reason with generics with an eye towards developing inheritance proof theory accordingly. This seems a much more sound methodology than relying solely on logicians' introspections about what is reasonable to conclude under particular interpretations of an inheritance network. For example, a recurrent theme in the thesis is that the status of 'redundant' links should not be taken for granted. It is a matter for speculation how the fruits of such research would reintegrate into the applications of inheritance reasoning and default reasoning in general that were outlined in the introductory chapter, but certainly it cannot hurt existing practical applications if the result includes more efficient reasoning strategies, and it can only benefit theoretical applications that invoke inheritance if a 'more correct' or 'best-fit' system is identified.

Appendix A

Precis of Veltman's "Defaults in Update Semantics"

A.1 Introduction

This appendix provides a technical summary of Veltman's work in dynamic semantics (Klein et al., 1990; Veltman, 1994), which aims to provide a formal framework in which to understand the semantics of defaults and natural language generics. Veltman (1994) develops a complex system in an incremental fashion starting from a framework similar to that reported by Klein et al. (1990) for an analysis of defaults in terms of their dynamic effect as sequences of sentences applied to information states. This is a quite general framework in which updates, tests, and revisions ('downdates') can ultimately be accommodated, although the initial work in this thread of dynamic semantics has postponed consideration of revisions. The gist of the approach is to model states by sets of atomic sentences, starting with the powerset of possible sentences, and eliminating those subsets that are inconsistent with input sentences as processing moves along.

A.2 Basic Definitions

Definition 69 *Let W be the universe, the powerset of a set of atomic sentences. An* expectation pattern *on W is a preordering ϵ of W: ϵ is reflexive and transitive. $\langle w, v\rangle \in \epsilon \equiv w \preceq_\epsilon v$.*

Essentially, an expectation pattern does the work for Veltman that preferred models do for Boutilier or that the accessibility relation among worlds does for Delgrande. However, Veltman's is a more finely grained system in that the pattern can change from state to state, and within a state there can be a multiplicity of expectation patterns (because π is a function that assigns patterns to subsets of sentences and not a single pattern to all subsets) so that different expectations can be brought to bear depending on the subset of information under consideration.

Definition 70 *Let ϵ be a pattern on W;*

- *w is a* normal world *in ϵ iff $w \in W$ and $w \preceq_\epsilon v$ for every $v \in W$;*
- *$\mathbf{n}\epsilon$ is the set of all normal worlds in ϵ*
- *ϵ is* coherent *iff $\mathbf{n}\epsilon \neq \emptyset$.*

Definition 71 *Let ϵ be a pattern on W and $s \subseteq W$.*

- *w is* optimal *in $\langle \epsilon, s\rangle$ iff $w \in s$ and there is no $v \in s$ such that $v \prec_\epsilon w$.*
- *$\mathbf{m}_{\langle \epsilon, s\rangle}$ is the set of all optimal worlds in $\langle \epsilon, s\rangle$.*

Definition 72 *Let ϵ and ϵ' be patterns on W, and $e \subseteq W$.*

1. *ϵ' is a* refinement *of ϵ iff $\epsilon' \subseteq \epsilon$;*
2. *$\epsilon \circ e = \{ \langle v, w\rangle \in \epsilon |$ if $w \in e$, then $v \in e\}$; $\epsilon \circ e$ is the* refinement of ϵ with the proposition *e.*

Given that the refinement operator has such an important role in updating information states, it is useful to illustrate its function with an example. Consider the following atomic sentences: a, b, c.

(A.1) $$W = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

Enumerate the elements of W using the position in the ordering as a name for each of the worlds; $w_0 = \emptyset$, $w_1 = \{a\}$ and so on. Let ϵ be a pattern on W, the reflexive and transitive closure of the following relation:

$\{ \langle w_0, w_1\rangle, \langle w_0, w_2\rangle, \langle w_0, w_3\rangle, \langle w_1, w_4\rangle, \langle w_1, w_6\rangle, \langle w_2, w_4\rangle, \langle w_2, w_5\rangle, \langle w_3, w_6\rangle, \langle w_3, w_5\rangle, \langle w_4, w_7\rangle, \langle w_6, w_7\rangle, \langle w_5, w_7\rangle \}$,

depicted graphically in Figure A.1. Also let $e = \{w_2, w_4, w_7\}$ and $e' = \{w_2, w_6, w_7\}$.

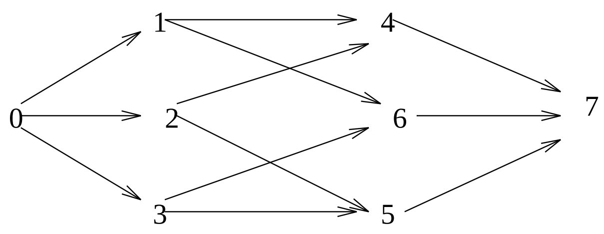


Figure A.1: A Pattern on W

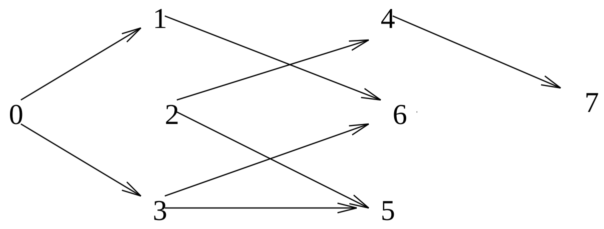


Figure A.2: A Refinement of a Pattern on W

Then, $\epsilon \circ e$ is the closure of the relation depicted in Figure A.2. There are no normal worlds in the refinement since $w_0 \not\leq_{\epsilon \circ e} w_2$, but w_2 is optimal in e. Figure A.3 depicts $\epsilon \circ e'$; again, there are no normal worlds, but w_2 and w_6 are optimal.

Ultimately, a number of patterns will be part of a complex function π that is part of an information state (the other part of the information state being a subset of W).

Definition 73 *Let W be the universe.*

- *A* frame on W *is a function π that assigns to every subset d of W a pattern πd on d.*

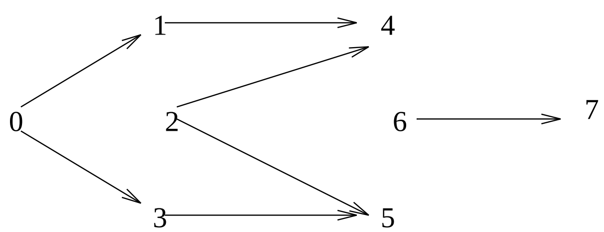


Figure A.3: Another Refinement on ϵ

- *Let π be a frame on W and $d, e \subseteq W$. The proposition e is a default in πd iff $d \cap e \neq \emptyset$ and $\pi d \circ e = \pi d$.*

In the above example (Figures A.1, A.2 and A.3), clearly, neither e nor e' is a default in ϵ.

Definition 74

Given a frame π on W, π is coherent *iff for every nonempty $d \subset W$, $\mathbf{n}\pi d \neq \emptyset$.*

With these preliminary definitions in place it is possible to formalize the idea of an information state.

Definition 75 *Let W be a model of the unverse, the powerset of the set of atomic sentences.*

1. *σ is an* information state *iff $\sigma = \langle \pi, s \rangle$, and one of the following conditions is fulfilled:*

   (a) π is a coherent frame on W, and s is a nonempty subset of W.

   (b) π is the frame $\langle \tau, \emptyset \rangle$, where $\tau d = \{\langle w, w \rangle | w \in d\}$ for all $d \subseteq W$.

2. *$\sigma = \langle \pi, s \rangle$* is at least as strong as *$\sigma = \langle \pi', s' \rangle$ iff $s \subseteq s'$ and $\pi d \subseteq \pi' d'$ for every $d \subseteq W$.*

3. *$\top$, the* minimal *state, is the state given by $\langle \upsilon, W \rangle$, where $\upsilon d = d \times d$ for every $d \subseteq W$.*

   $-$, the absurd *state, is the state $\langle \iota, \emptyset \rangle$.*

Although Veltman (1994) does not define the concept of a coherent information state, it seems that his intention is as follows:

Definition 76 *$\sigma = \langle \pi, s \rangle$ is a* coherent information state *iff σ is an information state where π is a coherent frame on W, and s is a nonempty subset of W.*

Veltman (1994, p. 34) defines:

Definition 77 *Let $e_1, \ldots, e_n$ be defaults in $\pi d_1, \ldots, \pi d_n$. A world w* complies *with $\{e_1, \ldots, e_n\}$ iff $w \in e_i$ for every i such that $w \in d_i (1 \leq i \leq n)$.*

Definition 78 *Let $\sigma = \langle \pi, s \rangle$ be a coherent information state and assume that $e_1, \ldots, e_n$ are defaults in $\pi d_1, \ldots, \pi d_n$, respectively.*

The set of defaults $\{e_1, \ldots, e_n\}$ applies within *s iff there is no $d \supseteq s$ such that $\mathbf{n}\pi d \subseteq \bigcup_{1 \leq i < n} (d_i \sim e_i)$.*

Instead of saying 'the set $\{e_1, \ldots, e_n\}$ applies within s', we often say '$e_1, \ldots, e_n$ jointly apply within *s.'*

Veltman relates definitions 77 and 78 by giving a reformulated version of the definition of the notion of a set of defaults applying within an information state in terms of a world's complying with each default. Here I make the relationship is more explicit by detailing in Proposition 14.

Proposition 14 *Let $\sigma = \langle \pi, s \rangle$ be a coherent information state. The set of defaults $\{e_1, \ldots, e_n\}$* applies within *s iff for every $d \supseteq s$ there is some $w \in \mathbf{n}\pi d$ such that w complies with $\{e_1, \ldots, e_n\}$.*

Proof: Let $\sigma = \langle \pi, s \rangle$ be a coherent information state.

$\Leftarrow$ Suppose that for every $d, s \subseteq d$ there is some w in $\mathbf{n}\pi d$ such that w complies with $\{e_1, \ldots, e_n\}$.
Then, for every $d, s \subseteq d$ there exists a $w \in \mathbf{n}\pi d$ such that $\forall i, w \in d_i \wedge w \in e_i$. $\therefore$ $\mathbf{n}\pi d$ contains elements that render it not a subset of $\bigcup_{1 \leq i < n} (d_i \sim e_i)$.

$\Rightarrow$ Suppose that the set of defaults $\{e_1, \ldots, e_n\}$ applies within s.
Then there does not exist a d at least as large as s such that $\mathbf{n}\pi d \subseteq \bigcup_{1 \leq i < n} (d_i \sim e_i)$. So, $\forall d, s \subseteq d, \exists w' \in \mathbf{n}\pi d$ such that $w' \notin \bigcup_{1 \leq i < n} (d_i \sim e_i)$, which means that $w' \in d_i \wedge w' \in e_i$.

$\square$

This just means that a default in the domain d applies within its subdomain s only when it applies in all the intermediate subdomains between s and d.

Definition 79 *Let $\sigma = \langle \pi, s \rangle$ be a coherent information state and assume that $e_1, \ldots, e_n$ are defaults in $\pi d_1, \ldots, \pi d_n$.*

1. *Then* $\{e_1, \ldots, e_n\}$ *is a* maximal applicable *set in* σ *iff* $e_1, \ldots, e_n$ *jointly apply within* s, *and for every* e_{n+1} *and* d_{n+1} *such that* e_{n+1} *is a default in* πd_{n+1}, *and* $e_1, \ldots, e_n, e_{n+1}$ *jointly apply within* s *it holds that* $e_{n+1} = e_i$, *and* $d_{n+1} = d_i$ *for some* $i \leq n$.
2. *A world* w *is* optimal in σ *iff* $w \in s$ *and* w *complies with a maximal applicable set of defaults. The set of optimal worlds is denoted by* $\mathbf{m}_\sigma$.

The whole point of all this is to give the machinery that specifies how to move from one state to another on the basis of information obtained through the interpretation of sentences that either offer additional facts, default rules, or tests of plausible conclusions. One more set of definitions (p. 32) is useful to facilitate the discussion of the actual interpretation of sentences.

Definition 80 *Let* W *be as before.*

1. σ *is an* information state *iff* $\sigma = \langle \pi, s \rangle$, *and one of the following conditions is fulfilled:*
   (a) π *is a coherent frame on* W, *and* s *is a nonempty subset of* W.
   (b) π *is the frame* $\langle \tau, \emptyset \rangle$, *where* $\tau d = \{\langle w, w \rangle | w \in d\}$ *for all* $d \subseteq W$.
2. $\sigma = \langle \pi, s \rangle$ is at least as strong as $\sigma = \langle \pi', s' \rangle$ *iff* $s \subseteq s'$ *and* $\pi d \subseteq \pi' d'$ *for every* $d \subseteq W$.
3. $\top$, *the* minimal *state, is the state given by* $\langle \upsilon, W \rangle$, *where* $\upsilon d = d \times d$ *for every* $d \subseteq W$.
   $-$, *the* absurd *state, is the state* $\langle \iota, \emptyset \rangle$.

Since updates change the information in states, a sequence of sentences applied in successive order to the minimal information state will yield states containing distinct domains and sets of expectation patterns on them.

A.3 Interpreting Defaults

The machinery described in the preceding section gets put to use during the interpretation of sentences like *normally* ψ and *presumably* ϕ in addition to basic propositional sentences such as θ.

Definition 81 *Let $\mathcal{A}$ be a set consisting of finitely many atomic sentences. Associate a language with $\mathcal{A}$, $L_0^{\mathcal{A}}$ that has $\mathcal{A}$ as its nonlogical vocabulary and for its logical vocabulary the following (classical) operators: $\neg, \vee, \wedge$ (as well as parentheses).*

Let $L_\delta^{\mathcal{A}}$ contain $L_0^{\mathcal{A}}$, along with the additional binary operator $\rightsquigarrow$ and unary operator presumably*.*

A formula ϕ is a sentence of $L_\delta^{\mathcal{A}}$ iff there are sentences ψ and χ of $L_0^{\mathcal{A}}$ such that $\phi = \psi$, or $\phi = \psi \rightsquigarrow \chi$, or $\phi =$ presumably *ψ.*

Asserted defaults are thus expressed using the binary connective: *if ψ then normally ϕ* is translated as $\psi \rightsquigarrow \phi$. Contingent propositions (*normally χ*) are expressed as $\chi \vee \neg\chi \rightsquigarrow \chi$. The sentence $\psi \rightsquigarrow \phi$ means that the proposition ϕ is a default in the domain of worlds picked out by ψ. Propositions from $L_0^{\mathcal{A}}$ are interpreted as follows:

Definition 82 *Let $\sigma = \langle \pi, s\rangle$ be an information state. For every sentence ϕ in $L_0^{\mathcal{A}}$, $\sigma[\phi]$ is determined as follows:*

- *if $s \cap [\![\phi]\!] = \emptyset$ then $\sigma[\phi] = -$,*
- *otherwise, $\sigma[\phi] = \langle \pi, s \cap [\![\phi]\!]\rangle$.*

The machinery introduced in the preceding section for refining expectation patterns is invoked in this context to identify the right domains.

Definition 83 *Let $\sigma = \langle \pi, s\rangle$ be an information state. $\sigma[\phi \rightsquigarrow \psi]$ is determined as follows:*

- *$\sigma[\phi \rightsquigarrow \psi] = -$ if $[\![\phi]\!] \cap [\![\psi]\!] = \emptyset$ or $\pi_{[\![\phi]\!] \circ [\![\psi]\!]}$ is an incoherent frame.*
- *Otherwise, $\sigma[\phi \rightsquigarrow \psi] = \langle \pi_{[\![\phi]\!] \circ [\![\psi]\!]}, s\rangle$*

These definitions detail the updates. Basic propositions refine the set of worlds consistent with the known description, without changing the expectation patterns inside the frame. Defaults refine the expectation patterns without refining the set of worlds. The nonupdating test associated with a proposition like *presumably ψ* is specified in Definition 84 at the outset.

Definition 84 *Let $\sigma = \langle \pi, s\rangle$ be a coherent information state $\sigma[$*presumably *$\psi]$ is determined as follows:*

- *If* $\mathbf{m}_\sigma \cap [\![\psi]\!] = \mathbf{m}_\sigma$, *then* σ[presumably $\psi] = \sigma$.
- *Otherwise,* σ[presumably $\psi] = -$.

Default conclusions are *tests* in update semantics; while in other systems one concludes ϕ by default, in update semantics one accepts the sentence *presumably* ϕ. This is the highest level definition that says when you can conclude *presumably* ϕ—if the most normal worlds with respect to the information state σ are exactly the worlds that contain ϕ (where worlds are modeled as sets of sentences), then the sentence *presumably* ϕ is true, and otherwise it is absurd. This is a dynamical notion of 'truth' at work since the definition actually states that if the antecedent condition holds then $\sigma[$ *presumably* $\phi] = \sigma$, which means that *presumably* ϕ is a sentence that performs a test on an information state such that if it's successful, you know that the information state supports the sentence, and otherwise it doesn't.

The significance of this fact is important to emphasize: it means that default conclusions are different in kind from the default rules that license them.[1] Simple facts and default rules, when added to information states *refine* those states, but in update semantics, the default conclusion is a test of what is reasonable to conclude on the basis of an information state. It is not something that refines the information state. This is quite different from the main body of literature in inheritance reasoning in which a great deal of the complications that are emerge relate to the implications of determining the closure of an inheritance network by including in it some of the implicit links. This is, in fact, the problem of stability. Update semantics takes a rather pragmatic approach to the problem since the most obvious solution is indeed to differentiate that which you conclude by default because you know the default and that which you conclude by chaining defaults.

[1]Also, note that a default here isn't a default rule $a \rightsquigarrow b$; rather, a default is a contingent proposition.

Bibliography

Aït-Kaci, H. (1984). *A Lattice Theoretic Approach to Computation Based on a Calculus of Partially Ordered Types.* Ph.D. thesis, Computer and Information Sciences, University of Pennsylvania.

Al-Asady, R. & Narayanan, A. (1993). More Notes on 'A Clash of Intuitions'. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, pp. 682–7. Chambery, France.

Bacchus, F. (1989). A Modest, but Semantically Well Founded, Inheritance Reasoner. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 1104–9. Detroit, Michigan.

Ballim, A., de Ram, S. C., & Fass, D. (1989). Reasoning Using Inheritance from a Mixture of Knowledge and Beliefs. In *Knowledge Based Computer Systems, Proceedings of KBCS89 (Bombay, India)*, pp. 387–96. Lecture Notes in Artificial Intelligence 444. Berlin: Springer Verlag.

Barwise, J. (1989). Conditionals and Conditional Information. In *The Situation in Logic*, pp. 97–135. Centre for the Study of Language and Information, Stanford University, CA. CSLI Lecture Notes Number 14; originally appeared in Traugott, et al. (eds), *On Conditionals*, Cambridge University Press (1986).

Barwise, J. (1991). Information Links in Domain Theory. Tech. rep. IULG-91-7, Indiana University Logic Group.

Barwise, J. (1993). Constraints, Channels, and the Flow of Information. In Aczel, P., Israel, D., Katagiri, Y., & Peters, S. (Eds.), *Situation Theory and its Applications, Volume 3*, pp. 3–27. CSLI Lecture Notes no. 37.

Barwise, J. & Etchemendy, J. (1987). *The Liar: An Essay on Truth and Circularity.* New York: Oxford University Press.

Barwise, J. & Seligman, J. (1993). Imperfect Information Flow. In *Proceedings of the Eighth IEEE Symposium on Logic in Computer Science.* Washington, D.C., IEEE Computer Society.

Barwise, J. & Seligman, J. (1994). The Rights and Wrongs of Natural Regularity. In Tomberlin, J. (Ed.), *Philosophical Perspectives, Vol 8.* California: Ridgeview.

Boutilier, C. (1989). A Semantical Approach to Stable Inheritance Reasoning. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 1134–39. Detroit, Michigan.

Boutilier, C. (1991). Inaccessible Worlds and Irrelevance: Preliminary Report. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pp. 413–18. Sydney, Australia.

Boutilier, C. (1993). On the Semantics of Stable Inheritance Reasoning. *Computational Intelligence*, *9*(1), 73–110.

Brachman, R. J. (1983). What ISA Is and Isn't: An Analysis of Taxonomic Links in Semantic Networks. *IEEE Computer*, *16*(10), 30–6.

Brachman, R. J. (1985). I Lied about the Trees, or Defaults and Definitions in Knowledge Representation. *The AI Magazine*, *6*, 80–93.

Braisby, N., Franks, B., & Hampton, J. (1994). On the Psychological Basis for Rigid Designation. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 56–65. Atlanta, Georgia.

Calder, J. (1989). Paradigmatic Morphology. In *Proceedings of the 4th Annual Meeting of the EACL*, pp. 58–65. Manchester, England.

Calder, J. (1990). *An Interpretation of Paradigmatic Morphology.* Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Carpenter, B. (1992). *The Logic of Typed Feature Structures.* Cambridge Tracts in Theoretical Computer Science 32, Cambridge: Cambridge University Press.

Cavedon, L. (1992). Conditional Constraints and Default Reasoning. Unpublished Manuscript. Centre for Cognitive Science, University of Edinburgh.

Cavedon, L. (1993). Reasoning about Events: The Qualification Problem. Unpublished manuscript. Centre for Cognitive Science, University of Edinburgh.

Cavedon, L. (1995). *A Channel Theoretic Approach to Conditional Reasoning.* Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Cavedon, L. & Glasbey, S. (1994). An Information-Flow Model of Generics. Presented at the Fifth Symposium on Logic and Language, Nosvaj, Hungary, September 2–6 1994. Submitted to *Acta Linguistica*.

Chater, N. & Oaksford, M. (1990). Autonomy, Implementation and Cognitive Architecture: A Reply to Fodor and Pylyshyn. *Cognition*, *34*, 93–107.

Cripps, A. (1987). Aspects of the Inheritance Problem. MSc Thesis, Department of Computer Science, University of Victoria.

Daelemans, W., Smedt, K. D., & Gazdar, G. (1992). Inheritance in Natural Language Processing. *Computational Linguistics*, *18*(2), 205–218. Special Issue on Inheritance: I.

Delgrande, J. (1987). A First-Order Conditional Logic for Prototypical Properties. *Artificial Intelligence*, *33*(1), 105–30.

Delgrande, J. (1988). An Approach to Default Reasoning Based on a First-Order Conditional Logic: Revised Report. *Artificial Intelligence*, *36*(1), 63–90.

Delgrande, J. (1990). A Semantics for a Class of Inheritance Networks. In *Proceedings of the 8th Biennial Conference of the Canadian Society for Computational Studies of Intelligence*. Ottawa, Ontario.

Dimopoulos, Y. (1992). On the Semantics of Inheritance Networks. *Journal of Experimental and Theoretical Artificial Intelligence*, *4*(1), 19–28.

Dorosh, J. & Loui, R. P. (Eds.). (1989). *The Workshop on Defeasible Reasoning with Specificity and Multiple Inheritance*. Washington University, St. Louis, MO. April, 1989. Workshop Problems.

Elio, R. & Pelletier, F. J. (1993). Human Benchmarks on AI's Benchmark Problems. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, pp. 406–411. June 18-21, 1993. Boulder, Colorado.

Elio, R. & Pelletier, F. J. (1994). The Effect of Syntactic Form on Simple Belief Revisions and Updates. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 260–5. Atlanta, Georgia.

Evans, R. & Gazdar, G. (1989). Inference in DATR. In *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 66–71. Manchester, England.

Garey, M. R. & Johnson, D. S. (1979). *Computers and Intractibility: A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman and Co.

Geffner, H. & Verma, T. (1989). Inheritance = Chaining + Defeat. In *CSD-890039*. Computer Science Department, University of California, Los Angeles.

Genesereth, M. R. & Nilsson, N. J. (1988). *Logical Foundations of Artificial Intelligence*. Palo Alto, CA: Morgan Kaufmann.

Goldman, Alvin, I. (1986). *Epistemology and Cognition*. Cambridge: Harvard University Press.

Hahn, U. & Vogel, C. (1995). Notes on Lawgicisim: Formalization, Computation, and Law. Presented at the *17th IVR '95 World Congress*, Bologna, Italy, June 16-20, 1995.

Healey, P. & Vogel, C. (1994a). Distributed Content in Natural Language Semantics. In *Third Conference of the European Society for Philosophy and Psychology*. Paris, France.

Healey, P. & Vogel, C. (1994b). A Situation Theoretic Model of Dialogue. In Jokinen, K. (Ed.), *Pragmatics in Dialogue Management*, Vol. 71, pp. 61–79. Gothenburg Papers in Theoretical Linguistics. Proceedings of the Special Session within the XIVth Scandanavian Conference of Linguistics and VIIIth Conference of Nordic Linguistics. August 16-21, 1993.

Henle, M. (1962). On the Relation between Logic and Thinking. *Psychological Review*, *69*, 366–78.

Hewson, C. & Vogel, C. (1994). Psychological Evidence for Assumptions of Path-Based Inheritance Reasoning. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 409–14. Atlanta, Georgia.

Horty, J. (1989). Discussion Session: Research Strategies: Individual Perspectives. In Dorosh, J. & Ronald P. Loui, e. (Eds.), *Edited Transcription of The Workshop on Defeasible Reasoning with Specificity and Multiple Inheritance, St.*

Louis, April 1989. Washington University, St. Louis, MO. Panel Members: David Poole, Fahiem Bacchus, James Delgrande, John Horty; Ben Grosof, Moderator.

Horty, J. & Thomason, R. (1988). Mixing Strict and Defeasible Inheritance. In *Seventh National Conference on AI*, pp. 427–32. AAAI-88, St Paul, Minnesota.

Horty, J., Thomason, R., & Touretzky, D. (1990). A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks. *Artificial Intelligence*, *42*(2-3), 311–48.

Kaufmann, H. & Goldstein, S. (1967). The Effects of Emotional Value of Conclusions upon Distortions in Syllogistic Reasoning. *Psychonomic Science*, *7*, 367–8.

Kautz, H. A. & Selman, B. (1991). Hard Problems for Simple Default Logics. *Artificial Intelligence*, *49*, 243–79.

Klein, E., Moens, M., & Veltman, F. (1990). Default Reasoning and Dynamic Interpretation of Natural Language. In van der Hoek, W., Meyer, J.-J. C., Tan, Y., & Witteven, C. (Eds.), *Non-Monotonic Reasoning and Partial Semantics*, pp. 21–36. New York: Ellis Horwood.

Lascarides, A., Briscoe, T., Asher, N., & Copestake, A. (1995). Order Independent and Persistent Typed Default Unification. In Schöter, A. & Vogel, C. (Eds.), *Nonclassical Feature Systems*, Vol. 10, pp. 61–135. An early version of a paper to appear in *Linguistics and Philosophy.*

Lascarides, A. & Copestake, A. (1995). Yet Another Default Unification. Unpublished manuscript, Centre for Cognitive Science, University of Edinburgh.

Lascarides, A. & Oberlander, J. (1993). Temporal Coherence and Defeasible Knowledge. *Theoretical Linguistics*, *19*, 1–37.

Laurent, D. & Vogel, C. (1994). Research Methodologies in the Context of Massive Electronic Connectivity. Presented at *Computers in Psychology*, York, England. September 1994. Available as Research Paper EUCCS/RP-67, Centre for Cognitive Science, The University of Edinburgh.

Lemon, O. J. (1995). First-Order Theory Change Systems with Intuitionist and Dynamic Semantics. In *Proceedings of Logic'95, 9th International Colloquium.* Liblice Castle, Czech Republic, June 13–16, 1995.

Lifschitz, V. (1989). 25 Benchmark Problems in Nonmonotonic Reasoning. In Reinfrank, M., de Kleer, J., & Ginsberg, M. (Eds.), *Workshop on Non-Monotonic Reasoning.* Berlin: Springer. Lecture Notes in Artificial Intelligence 346.

Maugham, S. (1919). *The Moon and Sixpence.* London: Pan Books Ltd. Pan Edition, 1974.

McDermott, D. & Doyle, J. (1980). Non-Monotonic Logic I. *Artificial Intelligence*, *13*(1-2), 41–72.

Morgan, J. J. B. & Morton, J. T. (1944). The Distortion of Syllogistic Reasoning Produced by Personal Convictions. *Journal of Social Psychology*, *20*, 39–59.

Niemelä, I. & Rintanen, J. (1994). On the Impact of Stratification on the Complexity of Nonmonotonic Reasoning. *Journal of Applied Non-Classical Logics*, *4*(2), 141–79.

Pearl, J. (1987). Probabilistic Semantics for Inheritance Hierarchies with Exceptions. In *TR-93*. Cognitive Systems Laboratory, University of California, Los Angeles.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.

Pelletier, F. J. & Elio, R. (1993). Some Truths about Non-Monotonic Reasoning. Tech. rep. TR 93-12, Department of Computing Science, University of Alberta, Canada.

Pollard, C. & Sag, I. (1987). *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. Centre for the Study of Language and Information, Stanford University, CA.

Pollard, C. & Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press and CSLI Publications.

Przymusinski, T. C. (1989). Three-Valued Formalizations of Non-Monotonic Reasoning and Logic Programming. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pp. 341–8. Toronto, Canada.

Reiter, R. (1980). A Logic for Default Reasoning. *Artificial Intelligence*, *13*, 81–132.

Reiter, R. & Criscuolo, G. (1983). Some Representational Issues in Default Reasoning. *Computers and Mathematics with Applications*, *9*(1), 15–27.

Rips, L. & Marcus, S. (1977). Supposition and the Analysis of Conditional Sentences. In Just, M. & Carpenter, P. A. (Eds.), *Cognitive Processes in Comprehension*. Hillsdale, NJ: Erlbaum.

Sandewall, E. (1986). Nonmonotonic Inference Rules for Multiple Inheritance with Exceptions. *Proceedings of IEEE*, *74*(10), 1345–53.

Schubert, L. & Pelletier, F. J. (1987). Problems in the Representation of the Logical Form of Generics, Plurals, and Mass Nouns. In LePore, E. (Ed.), *New Directions in Semantics*, pp. 385–451. London: Academic Press.

Seligman, J. (1990). Perspectives in Situation Theory. In Cooper, R., Mukai, K., & Perry, J. (Eds.), *Situation Theory and its Applications, Volume I*, pp. 147–191. Centre for the Study of Language and Information, Stanford University, CA.

Seligman, J. & Barwise, J. (1993). Channel Theory: Toward a Mathematics of Imperfect Information Flow. Unpublished manuscript.

Selman, B. & Levesque, H. (1989). The Tractability of Path-Based Inheritance. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 102–9. Detroit, Michigan.

Selman, B. & Levesque, H. (1993). The Complexity of Path-Based Defeasible Inheritance. *Artificial Intelligence*, *62*, 303–39.

Stein, L. A. (1989). Skeptical Inheritance: Computing the Intersection of Credulous Extensions. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 1153–60. Detroit, Michigan.

Taplin, J. E. & Staudenmayer, H. (1973). Interpretation of Abstract Conditional Sentences in Deductive Reasoning. *Journal of Verbal Learning and Verbal Behavior*, *12*, 530–42.

Touretzky, D. (1986). *The Mathematics of Inheritance Systems.* Los Altos, CA: Morgan Kaufman.

Touretzky, D., Horty, J., & Thomason, R. (1987). A Clash of Intuitions: The Current State of Non-Monotonic Inheritance Systems. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pp. 476–82. Milan, Italy.

Veltman, F. (1991). Defaults in Update Semantics. Tech. rep. LP-91-02, Institute for Language, Logic and Information, University of Amsterdam.

Veltman, F. (1994). Defaults in Update Semantics. Unpublished ms.; Institute for Language, Logic and Information, University of Amsterdam.

Vogel, C. (1992). A Situation Theoretic Model of Inheritance Reasoning. In *Logic, Language, and Information: Toward an Integrated Theory of Linguistic Processing*. European Science Foundation. Unpublished Manuscript presented at the ESF Research Conference. Autrans, France, December 13-17, 1992.

Vogel, C. & Cooper, R. (1994). Robust Chart Parsing with Mildly Inconsistent Feature Structures. In AAAI Technical Report: AAAI 1994 Fall Symposium—*Knowledge Representation for Natural Language in Implemented Systems*. November 4-6, 1994, New Orleans, LA. A revised version appears in Andreas Schöter and Carl Vogel (eds.) *Nonclassical Feature Systems*. Edinburgh Working Papers in Cognitive Science, Volume 10. (1995) EUCCS-WP10. pp. 197-216.

Vogel, C., Popowich, F., & Cercone, N. (1993). Logic-Based Inheritance Reasoning. In *Proceedings of the 9th Bi-Annual Conference of the Society for the Study of Artificial Intelligence and the Simulation of Behavior*, pp. 179–88. AISB'93, Birmingham, England. March 29–April 2, 1993. Appears in, A. Sloman, et al. (eds.), *Prospects for Artificial Intelligence*. Amsterdam: IOS Press. 1993.

Wilkins, M. C. (1928). The Effect of Changed Material on Ability to Do Formal Syllogistic Reasoning. *Archives in Psychology*, no. 128.