Psychological Constraints on Plausible Default Inheritance Reasoning

Carl Vogel & Judith Tonhauser
Institute for Computational Linguistics
University of Stuttgart
Azenbergstr 12
D-70174 Stuttgart, Germany
{vogel,tonhauser}@ims.uni-stuttgart.de

Abstract

Human responses to a large number of problems that are relevant to the literature on default reasoning are examined. Two foundational assumptions of the literature are identified: chaining of defaults is possible; chaining is not possible beyond a negative default. These assumptions are used as *strict* criteria to select subjects who always behaved in accord with them. Their within-subject patterns of response are examined. The paper identifies 5 reasoning strategies: skepticism, explicit link acceptance, shortest path reasoning, most paths reasoning, and a combination of the latter two reasoning. Of these, the literature has hitherto regarded seriously only extreme skepticism; for example, shortest-path reasoning has been discounted since (Touretzky, 1986). The strict inclusion criteria are then relaxed, and over a space of consistent reasoners the same reasoning strategies remain useful in classifiying human response. If we want machines to reason about defaults the way people do, or reason about people who reason with defaults, the field must address these easily formalized modes of reasoning.

1 INTRODUCTION

Hierarchies are ubiquitous in native human organization of information and often contain tangles and defaults (e.g. conceptualizations of social/honorific hierarchies). Path-based default inheritance reasoning is interesting as a form of nonmonotonic inference with tangled hierarchies because of its attractive computational properties. Inheritance can be viewed as a variant of the monadic predicate calculus, with a guarantee of inferential acyclicity, and for this reason should always be decidable, sometimes tractable.[1] For example, the skeptical theory of inference proposed by Horty et al. (1990) is polynomial, and of course more restricted systems perform even better (Niemelä & Rintanen, 1994). However, efficiency is not always a property of inheritance reasoners (Selman & Levesque, 1989, 1993) and certainly not of nonmonotonic inference at large. Complexity is introduced in the varying ways that inference can be defined on the acyclic graphs isomorphic to inheritance theories. There are, of course, any number of ways of defining this inference, and competing methods appeal to 'intuitions' about which methods are most appropriate (Touretzky, Horty, & Thomason, 1987). The intuitions at stake are those which define the most plausible conclusions on the basis of particular interpretations of graphs for those graphs which different conclusions depending upon the reasoner applied. Arguments about plausibility, rather than validity, are made because inheritance is designed to be a psychologically plausible model of human reasoning with defaults.

Inheritance reasoners purport to provide a psychologically plausible model of reasoning with defaults, partially motivated by the ubiquity of tangled hierarchies themselves in the organization of information, and partially because the efficient decision procedures associated with the representation make it seem a reasonable descriptor of human reasoning, which is relatively efficient. 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 exceptional objects in light of default theories in relation to the way general default logics classify the same exceptional objects. They also present the first pilot study applying similar scrutiny to inheritance reasoners. Hewson and Vogel (1994) and Vogel (1995) directly test the plausibility of inheritance reasoning, in an attempt to assess the degree of fit that popular inheritance reasoners (like that of Horty et al. (1990); hereafter called H90) have with the data supplied by human reasoning. This paper describes an extension of that work.

[1] Thanks to Jeff Pelletier for making that clear to us.

The data from their experiments is pooled into a substantial body of evidence on human reasoning with sets of defaults. Rather than assessing the degree of fit of extant systems with the data (since none of the path based systems fit very well), observations are made concerning the features of reasoners which would provide a good predictor of human reasoning, and a set of reasoners derived from these observations are characterized. We proceed by outlining our assumptions and characterizing the experiments from which the data analyzed here are drawn. Then we define a consistency criterion with respect to satisfying transitivity and negative reasoning which we apply strictly to select those subjects who were most consistent in their overall replies, and outline the reasoning strategies which are predictive of their responses. We then provide a similar analysis for subjects grouped according to a range of degrees of satisfying the inclusion criterion. The same basic patterns are apparent at each stage: primarily full skepticism, and among the nonskeptics, a mixture of a kind of shortest-path/most-paths reasoning.

1.1 DEFINING PLAUSIBILITY

The psychological plausibility of a logic can be defined as the degree to which it captures the reasoning patterns that people ordinarily use. One can define logics of 'ideally rational agents'; however, those logics are nearly always undecidable, and therefore are far from 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. We are not immediately concerned with whether logics provide a correct morphism to the processes that actually govern human reasoning; rather, our interest is the sort of logic that provides the best description language for expressing exactly those inferred sentences that people are likely to agree are true or worth acting upon. If one's goal is to describe human reasoning, then this is a reasonable way of proceeding. If the goal is to develop a machine that reasons more correctly than humans, *if it is to interact with humans it will require 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 correspondences between proof theoretic parameters and reasoning strategies that people use (consistently or in different circumstances).

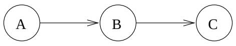
1.2 TESTING PLAUSIBILITY

Hewson and Vogel (1994) addressed inheritance reasoning directly by attempting to determine what sorts of conclusions people find in sets of abstract defaults. The statements in the sets were abstract in the sense that uninterpreted roman letters were used as labels for concepts (*Bs are normally Cs*) rather than given concrete interpretations (*Marines are normally short haired*), to allow control over the influence of other knowledge, belief or opinion (Kaufmann & Goldstein, 1967). Subjects were supplied five possible answers to each problem, with appropriate instantiations for X and Y depending on the exact problem: *Xs are normally Ys*; *Xs are normally not Ys*; *Xs are normally Ys and Xs are normally not Ys*, *it is indeterminate whether Xs are normally Ys or normally not Ys*, *I don't know*. The materials were questionnaires comprised of 40 problems, each problem containing a set of statements constituting a default theory about abstract concepts. Subjects were asked to give the answer they thought appropriate on the basis of the information given in each problem alone—it was stressed that there were no right or wrong answers, and subjects were to say what *they* thought could be concluded. Materials were supplied with the default theories in three modes of presentation: sentence only, graph only, sentence and graph. Each of 72 subjects (in a range of ages and generally with non-technical backgrounds) received all 40 questions (in a random order or the reverse) in one mode of presentation. The 40 problems were designed to encompass a number of inheritance networks specifically disputed in the literature in relation to the conclusions they should 'intuitively' license, the choice of answers constrained by the sorts of classifications of the problems made in the the literature. An example problem from the graph+sentence mode of presentation is given in Figure 1. The problems were comprised solely of defaults. In the graph-only condition, subjects were told only in the instructions that the arrow meant 'are normally', and thus had substantially less priming of 'defaultness' than subjects in sentential conditions. It is convenient to index each problem by its graph presentation. In each problem, the respondent is asked to characterize the relationship between the 'leftmost' and 'rightmost' class mentioned in the premises.

Some problems were set to test the more foundational assumptions of the inheritance literature: transitivity, negativity, redundancy and preemption. The remaining problems were controls. Hewson and Vogel (1994) reported surprising findings. Firstly, most of the problems were rated indeterminate.[2] Of the determinate response, while there was considerable support for transitivity, and some support for simple cases of preemption, there was little evidence for redundancy (redundant statements/links seemed to be interpreted as providing additional information rather than re-

[2]Answers in categories 'c' or 'd' are referred to as *indeterminate classification* of a problem. An answer in category 'a' is *positive classification*; 'b', *negative classification*. An answer of either 'a' or 'b' is a *definite classification* or *determinate*.

$\boxed{1}$



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

What can you conclude from these statements? 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 1: An Example Question

dundant and disregardable information), and people seemed to reason with negative information in a way wholly unpredicted by the inheritance literature (Vogel, 1996). Hewson and Vogel (1994) did not report evidence about the various specific forms of preemption or subpath ambiguity. There was no clear finding with respect to mode of presentation, except that graphic presentation seemed to polarize responses more, suggesting that the graphic syntax of inheritance reasoning has much to do with the strength of 'intuitions' that have been discussed in the literature. Vogel (1995) partially replicated the experiment of (Hewson & Vogel, 1994), dropping the graph+sentence condition and testing the materials on an additional 98 subjects (undergraduates in English literature and composition courses); the results corroborated the earlier findings.

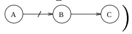
This paper pools the data from the 2 experiments. After removal of unattempted questionnaires, and those for which collation errors during material preparation resulted in different questionnaires than other subjects had (due to repetition & omission of sheets), there are 162 questionnaires in total. With 40 questions each there were potentially 6480 datapoints. However, altogether there were 71 unanswered problems, leaving 6409 datapoints in total. This is a substantial body of data which is open to a great deal more analysis than Hewson and Vogel (1994) or Vogel (1995) supplied. Here we will examine particular descriptors of response patterns that can be articulated as reasoning strategies. For discursive convenience, this paper will refer to the *descriptors* of reasoning patterns as if they were *strategies* that determined responses. Thus, this paper outlines three nonskeptical inference strategies that were consistently 'employed'. The strategies are identified in examining patterns of responses within subjects to sets of problems from the described experiments and sets the groundwork for examining their between subjects systematicity. These strategies are then taken as formal definitions of inheritance reasoners and the predictive efficacy of each will be evaluated.[3]

2 PSYCHOLOGICALLY PLAUSIBLE REASONERS

In this section we examine response patterns of subjects who satisfied criteria of regularity in responses to related problems. We initially adopt absolute consistency criteria relative to basic assumptions of inheritance reasoners and then examine the results among subjects who satisfied these criteria in gradual degrees. The hope is that by identifying such systematicity in response patterns we can identify or define reasoners which yield the closest approximation to observations. We consider within-subject patterns of reasoning (the previously mentioned analyses considered only between-subject systematicity for individual problems; the current work attempts to assess between subject systematicity over *sets* of related problems). There are two rational patterns of reply which we considered. The first was dictated by basic assumptions of the literature and the second by systematic behaviors which actually appeared. For both patterns, under the absolute criteria, data obtained from subjects who did not conform to predictions of positive and negative transitivity (measured by problems containing no conflicts) were eliminated; this left 88 questionnaires. This is motivated by the consideration that the reasoning of those subjects whose responses were not predicted by transitivity on graphs containing no conflicts would be better predicted by classical statistical models (and are thus outside the scope of default inheritance) or were otherwise relatively inconsistent in response patterns. The latter possibility is examined in § 2.2.

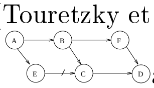
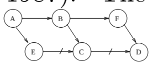
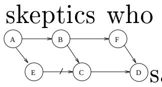
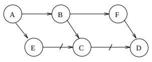
2.1 ABSOLUTE CONSISTENCY

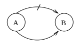
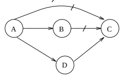
2.1.1 Negative Chains are Indeterminate

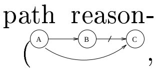
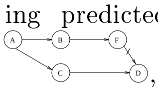
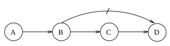
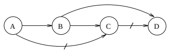
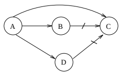
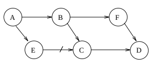
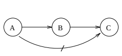
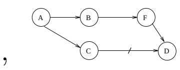
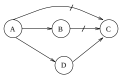
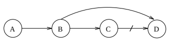
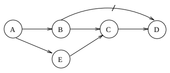
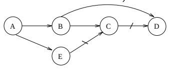
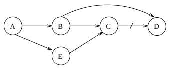
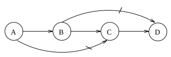
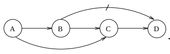
In the first absolute inclusion criteria, additional subjects whose responses to problems containing non-final negative links (e.g. ) propagated negativity (i.e., concluded definitely that *As* are normally not *Cs*)

[3]An informal aside on complexity: interesting (but inconclusive) facts about the time it took subjects to complete the experiments (we emphasize that these were not reaction-time experiments), conjoined with estimates on average reading time (e.g. Rayner & Pollatsek, 1989), suggest that linear reasoning methods (e.g. Niemelä & Rintanen, 1994) are too efficient and that polynomial methods (e.g. Horty et al., 1990) are just about right.

were also eliminated,[4] leaving the materials of 8 subjects. While this is a rather small resulting dataset, it should be pointed out eliminations were taken to be stringent: in the transitivity case, subjects were eliminated if they did not respond in accord with transitivity in each of the 8 graphs without conflicts; in the negativity case, subjects were eliminated if they propagated negativity with any of the 7 (different) graphs involving nonfinal negative links.[5] In this initial analysis we were less strict with respect to negativity cancellation than to propagation, in homage to the fact that humans find negative reasoning in general more difficult: if a subject gave a positive classification to at most one problem involving double negatives, we still admit them into this analysis. This resulted in the inclusion of one subject; thus, the inclusion criteria remain quite strict. A larger number of subjects satisfied the inclusion criteria to lesser degrees. In the context of the detailed within-subjects analysis, extreme stringency is most appropriate since it still illustrates the main points. Clear reasoning strategies emerge and are satisfied by other subjects as well. The eliminations made are both justified in that they leave only those subjects who are at least in accord with the foundational assumptions of default inheritance reasoning. The subsequent section (§ 2.2) examines subjects satisfying the inclusion criteria to varying degrees. The same basic reasoning patterns emerge among the additional 43 subjects as we find here among the 8.

Of the 8 subjects who satisfied the strict inclusion criteria, 3 were more or less credulous.[6] The remainder, with 2 exceptions, rated graphs as indeterminate unless there were no conflicting paths. In terms of the literature, they were *skeptical* in the face of conflicting defaults, yet still accepted chaining of uncontested defaults (which is classically invalid). The exceptions in this remainder are rather interesting, because they involved positive classification of 2 problems which are canonical for the dispute over whether ambiguity in a subpath should be cascaded (Touretzky et al., 1987). The graphs in question are  and . To the former, which the literature assumes to contain no conflicting paths because of the status of propagation of negativity, three of the 'skeptics' responded by giving it a positive classification: *As are normally Ds*. The other 2 skeptics classified the graph as indeterminate. The latter graph was classified positively by one of the skeptics, and as indeterminate by the other four. Curiously, none of the skeptics who determined that *As are normally Ds* of  said the same of . This means that the five skeptics were partitioned into: the one who classified both as indeterminate, the three who found the first one positive, and the one who found the second one positive. As far as adjudicating inheritance proof theory is concerned, this would argue for both forms of skepticism, with the restricted skepticism of Horty et al. (1990) fitting 25% of the time. Of the three 'credulous' reasoners, 2 classified both of the problems positively. Given that 'credulous' is being used in a rather nontechnical sense here (referring not to the sanctioning of a multiplicity of extensions, but to likelihood to make some consistent and definite classifications of problems) this actually means that there was equal support for both ambiguity propagating skepticism and the non-propagating restricted skepticism of Horty et al. (1990). No wonder there is a clash of intuitions.

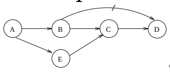
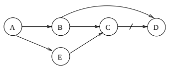
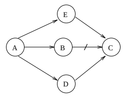
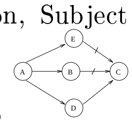
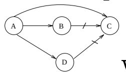
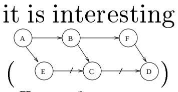
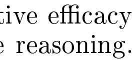
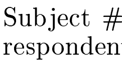
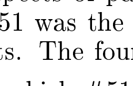
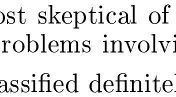
The 3 subjects who made more definite classifications than the skeptics supply interestingly consistent responses. Subject #62 was the most credulous. With 2 exceptions, the response strategy of #62 would most aptly be labeled *affirmation*: if a positive path existed #62 gave the problem a positive classification. The 2 exceptions are , which #62 classified as indeterminate ('c') and  (the nixon diamond with a direct negative link), which #62 classified as negative. However, the remaining 32 problems which did not involve a nonfinal negative link in a single chain of links were all given definite classifications, and the only other negative classifications among these 32 were of those problems in which there was no positive paths available.

Subject #12 was the next most credulous of the 3 in the sample under focus. Subject #12 favored the response of shortest path reasoning where conflicting paths existed and shortest path reasoning predicted a definite classification (, , , , , , , , , , , , , ). Shortest path reasoning is the best classification of this response strategy because,  was evaluated

[4]We refer to chains containing nonfinal negative links as *negative chains*.

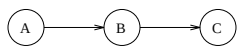
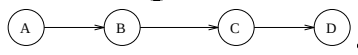
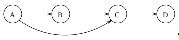
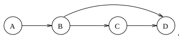
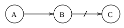
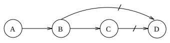
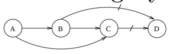
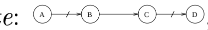
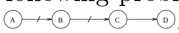
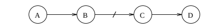
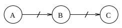
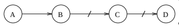
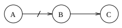
[5]If a chain contained two negative links, it would be possible for the negativity to either intensify, yielding a definite negative classification, or to cancel, yielding a definite positive classification.

[6]'Credulous' in the sense of the inheritance literature, means willingness to draw definite conclusions where others might be skeptical and draw none (e.g. faced with ambiguity).

by this individual as indeterminate, and this particular problem is the canonical example of why simple shortest path reasoning is not desirable (Touretzky, 1986), leading to the definition of more complex path preference strategies (which classify this problem negatively using answer 'b'). Subject #12 agreed with the shortest path prediction for this problem by classifying it as indeterminate. Interestingly, this pattern of classification also holds for 2 problems canonical for the dispute between on-path and off-path preemption:  and . Subject #12 classified both as a shortest path reasoner would. As it happens, this is the same classification predicted by off-path preempting reasoners like Horty et al. (1990) and Al-Asady and Narayanan (1993). This is interesting because off-path preempted chains can be seen as a species of 'redundant' links, and in the simple case this subject did not disregard the redundant information. However, in the simple case the 'redundant' link rendered the problem indeterminate with respect to shortest path reasoning. In response to problems which shortest path reasoning did not resolve to a definite solution, Subject #12 gave a *most paths* reply: ,  (note that  was classified with answer 'a'). Thus, Subject #12 seems to be representative of the following reasoning strategy: shortest path where that makes definite conclusions available, most paths where that makes definite conclusions available if shortest path reasoning doesn't, and indeterminate otherwise. Given this classification of the responses of Subject #12, it is interesting in addition that this subject classified () positive determinate, consistently with off-path preemption and non-cascaded ambiguities. Moreover, this subject was in a sentence only condition, thus exhibiting remarkably complex consistency for problems whose presentation did not directly support graphical classification; thus, this subject stands as an argument for the predictive efficacy of some aspects of path-based inheritance reasoning. Subject #51 was the most skeptical of the credulous respondents. The four problems involving conflicting arguments which #51 classified definitely (, ,  and ) were all classifiable by shortest path reasoning. While shortest path reasoning seems to be the simplest description of the strategy at work, it is interesting to note that in 3 of these cases, the shortest path was only one link long.

It is worth pointing out that the *affirmer* participated in the graph-only condition of the experiment, the *shortest path reasoner* participated in the sentence-only condition, and the *explicit link acceptor* (#51) was in the sentence+graph condition. Of the 5 skeptics, one was in the sentence-only condition; another, sentence+graph; the other 3, graph-only. This is an interesting distribution with respect to the distribution discussed in the next section and the overall inconclusive mode-of-presentation findings of Hewson and Vogel (1994). Only in the most consistent category are there so few subjects who were in the sentence-only condition. This is the clearest evidence to date that the conditions that offered graphical stimuli polarized response patterns. This is intuitively to be expected, since the sentence-only condition is quite difficult without visualizing the problems.

2.1.2 Negative Chains are Negative

In the last section we reported the within and between subjects response patterns among the 8 subjects whose behavior was most consistent, and which respected certain foundational assumptions of inheritance reasoning: we ruled out subjects who did not answer according to the predictions of transitivity in problems without conflicts, as well as those subjects who did not classify negative chains as indeterminate. For the first criterion subjects must have 'correctly' classified *each* of the 'valid' transitivity problems with *determinate* classifications, making positive definite classification of : , , , and , as well as answer category 'b' to each of, , , , and . In the case of negative chains, subjects must have classified *each* of the following problems as *indeterminate*: , , , , , , and . As we've seen, this reduced the pool of 162 subjects down to 8. In this section we examine the data admitted by an absolute relaxation of the negativity criterion. This is the approach taken by Vogel (1996) who found most paths reasoning to be a robust predictor.

First we describe the relaxation of the negativity criterion. It is a basic assumption within the inheritance literature that a negative link can occur only in a path-final position. A chain of links with a nonfinal negative link is a negative chain. In general, there is no valid information to be had about a definite relationship existing between the endpoints of such a chain. Thus, rating negative chains as conveying definite negative information is a different sort of divergence than labeling some cases of uncontested transitivities as indeterminate. However, it should be no surprise that negative reasoning is problematic, this result often appears in psychological investigation of human reasoning, for instance Wason and Johnson-Laird's (1972) card selec-

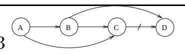
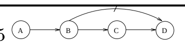
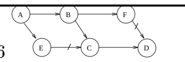
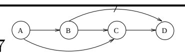
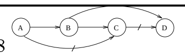
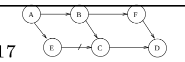
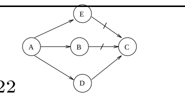
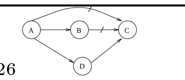
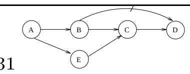
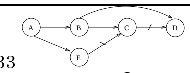
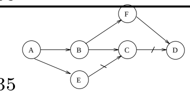
Problem		Response	
	a	b	c/d
 3	3	13	15
 5	12	2	18
 6	5	12	15
 7	15	3	14
 8	3	14	14
 17	15	1	16
 22	1	10	20
 26	1	13	18
 31	17	0	15
 33	1	11	19
 35	1	12	16

Table 1: Between-Subject Patterns of Response to Problems

tion task, and Evans' (1983) model construction task. The fact is that people are not very good at negative reasoning. Our conclusion, given that we think we should be interested in modeling human rationality, is that invalid negative reasoning involves rational, if pragmatically suspect, strategies. Given that we want to develop models of human reasoning, we should have a model which is inclusive of objective flaws in human reasoning.

Vogel (1996) analyzed the data using exactly this relaxation (*all* of the transitivity cases as predicted, and *all* of the negativity cases contrary to the literature's prediction) and thus obtained a pool of 32 subjects.[7] Table 1 depicts a between-subjects summary of responses to 11 problems for which significance obtained. Results for an additional 13 problems for which significance did not obtain are excluded here, one other problem which could have been used as an inclusion criterion was also excluded. Refer to Vogel (1996) for the full table. Table 1 shows that there is mainly indeterminacy, and in many cases a roughly equal amount of determinacy. For discursive convenience, call the problems with a preponderance of 'a' responses (5, 7, 17, and 31) among the determinate replies the *A-Problems*. The problems with significant numbers of 'b' classifications are 3, 6, 8, 22, 26, 33 and 35. Call those the *B-Problems*. Note that when negative chains are allowed to be classified as providing definite negative reasoning, then the pattern of reply on these problems fits with the predictions of most path reasoning.

A within-subject study of the data revealed that 11 subjects classified all of the A-Problems positively and all of the B-Problems negatively. That is, a full third of the subset were rather consistent internally. Another 3 subject did not classify all 4 of the A-problems positively, but did classify all of the B-problems negatively, and a further 6 subjects classified most of the A-problems affirmatively and tended to classify the B-problems negatively as well (3 or more). Thus, two-thirds of the subjects in this subset were in fact largely consistent with respect to the predictions of most paths reasoning in these cases. It is important to emphasize that the subjects were drawn from both the graphical conditions and the sentence-only conditions with the same patterns of reply in both of those conditions. Thus, it would appear that at least within this partition of subjects, with the allowance of negative chains as negative paths, most paths reasoning is the best predictor of replies. Note that the replies to problem 5 is counter to what is ordinarily predicted by the literature on path-based inheritance, as well as most paths reasoning.

2.2 DEGREES OF CONSISTENCY

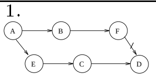
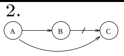
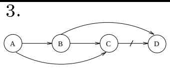
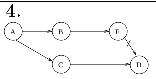
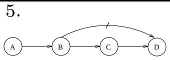
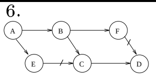
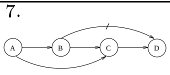
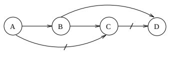
We have reported the within and between subjects response patterns among the subjects whose behavior was most consistent with respect to transitivity. We also considered two different absolute inclusion criteria regarding negative reasoning: the first (and most prolifically exclusive) was the assumption that negative chains are indeterminate; the second, that negative chains are negative. With a consistent pattern of reply, both patterns are rational. In the first case, the subject pool was reduced from 162 to 8, and in the second to 32. We reported the patterns of definite response to problems not among the inclusion criteria and identified potential reasoning strategies which would predict those patterns.

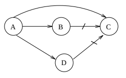
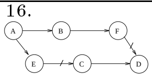
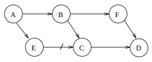
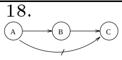
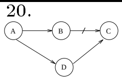
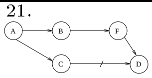
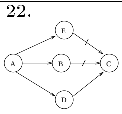
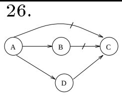
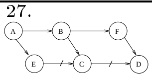
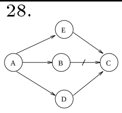
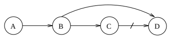
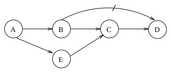
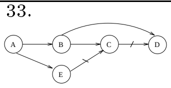
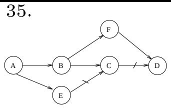
However, the patterns of response are in general systematic, even among subjects who were less consistent over the whole questionnaire. In this section we examine the data admitted by relaxing the inclusion criteria and demonstrating that the same patterns of response hold in the larger pool of subjects, albeit with increasing inconsistency of within-subject predictability corresponding to increasing inconsistency with respect to the inclusion criteria. There are three basic systematic strategies that can be used to do this. One approach is to hold fixed the idea that negative chains are indefinite and consider degrees of satisfaction of transitivity. Another is to hold transitivity fixed and

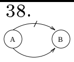
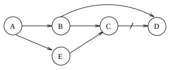
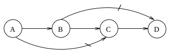
[7]By the definition of the inclusion criterion, these are exclusive of the 8 subjects discussed above.

consider increasing degrees of definite response to the negative chains (the extreme point of the latter relaxation yields exactly the subjects considered in the last section, and the minimally inconsistent along the former relaxation are exactly those subjects considered in §2.1.1). The third approach is to consider degrees of satisfaction on both axes of inclusion simultaneously.

At this point, since more problems will have definite answers as we progressively relax the inclusion criteria, it becomes useful to tabulate the 25 noninclusion problems along with the categories of response appropriate to the main reasoners we have addressed here: explicit link acceptance, shortest path reasoning, most paths reasoning. For comparison, we also give the answer to each problem according to H90 and to a reasoner just like H90 except more skeptical, using on-path preemption and not discounting paths with nonpermitted subpaths. The headings on the following list — **EL**, **SP**, **MP**, **H90** and **SK** — correspond to those reasoners. As we have also considered the possibility of negative chains being either indeterminate or negative, we mark responses made by a reasoner with an annotation indicating if its answer depends on negative chains being negative ($*^-$) or on them being indeterminate ($*_?$). The following table summarizes this information. Note that in some cases the reasoners are in agreement. In the analysis which follows we make within subject examinations of consistency in determining which strategies are most often representative.

Prob.	Resp.	EL	SP	MP	SK	H90
1. 	a					
	b					
	c/d	$*$	$*$	$*$	$*$	$*$
2. 	a	$*$	$*$		$*$	$*$
	b					
	c/d			$*$		
3. 	a				$*$	$*$
	b			$*$		
	c/d	$*$	$*$			
4. 	a		$*$			
	b					
	c/d	$*$		$*$	$*$	$*$
5. 	a					
	b		$*$		$*$	$*$
	c/d	$*$		$*$		
6. 	a					
	b			$*^-$		$*$
	c/d	$*$	$*$	$*_?$	$*$	
7. 	a			$*$		
	b				$*$	$*$
	c/d	$*$	$*$			
8. 	a		$*_?$		$*$	$*$
	b			$*^-$		
	c/d	$*$	$*^-$	$*_?$		

Prob.	Resp.	EL	SP	MP	SK	H90
9. 	a	$*$	$*$		$*$	$*$
	b			$*$		
	c/d					
16. 	a					
	b		$*$	$*$	$*$	$*$
	c/d	$*$				
17. 	a		$*_?$	$*$	$*$	$*$
	b					
	c/d	$*$	$*^-$			
18. 	a					
	b	$*$	$*$		$*$	$*$
	c/d			$*$		
20. 	a					
	b					
	c/d	$*$	$*$	$*$	$*$	$*$
21. 	a					
	b		$*$			
	c/d	$*$		$*$	$*$	$*$
22. 	a					
	b			$*$		
	c/d	$*$	$*$		$*$	$*$
26. 	a					
	b	$*$	$*$	$*$	$*$	$*$
	c/d					
27. 	a					$*$
	b			$*^-$		
	c/d	$*$	$*$	$*_?$	$*$	
28. 	a			$*$		
	b					
	c/d	$*$	$*$		$*$	$*$
29. 	a		$*$		$*$	$*$
	b					
	c/d	$*$		$*$		
31. 	a			$*$		
	b		$*$			$*$
	c/d	$*$			$*$	
33. 	a		$*$		$*$	$*$
	b			$*^-$		
	c/d	$*$		$*_?$		
35. 	a					$*$
	b			$*^-$		
	c/d	$*$	$*$	$*_?$	$*$	

Prob.	Resp.	EL	SP	MP	SK	H90
38. 	a					
	b					
	c/d	*	*	*	*	*
39. 	a		*			*
	b			*		
	c/d	*			*	
40. 	a					
	b		*	*$^{-}$	*	*
	c/d	*		*?		

Problem	Response a	b	c/d
3	0	1	3
5	0	3	1
6	1	1	2
7	0	0	4
8	3	0	1
17	3	0	1
22	0	1	3
26	0	4	0
31	0	3	1
33	1	1	2
35	0	0	0

Table 2: Transitivity Relaxed, Between-Subject Patterns of Response

2.2.1 Relaxing Transitivity

For the present analysis, we considered subjects whose response patterns were less consistent with respect to the transitivity requirements. We examined the patterns of response over the remaining problems which were given definite classifications. That is, we held fixed the negativity requirement from the stringent consistency criteria in which negative chains are indeterminate, but analyzed data from subjects who did not answer as predicted on all of the transitivity cases. This admitted another 8 subjects; however, inspection revealed that 3 of them did not get any of the transitivity cases as predicted, and 1 got only 2. Their materials were hard to find patterns in. Of the remainder, 3 subjects answered 6 of the transitivity problems according to the criteria and 1 of them answered 3 as predicted. It is those 4 subjects (#18, #54, #57, #81) whose patterns of reply we examine next.

Subject #81 was the most skeptical of these 4. This subject gave only 5 definite classifications among the 25 problems outwith the inclusion criteria. Of these, one classification was in accord with the only path in the network under the assumption that the negative chain was deemed indeterminate (an assumption that is, of course, consistent with the inclusion criteria assumed in this section). Another 2 problems were consistent with explicit link acceptance, and 2 more with shortest path reasoning. Thus, the behavior of this individual in problems classified definitely is best predicted by shortest path reasoning. However, we lack a classification of problems for which shortest path reasoning or explicit link acceptance could have been used, but wasn't, by this individual.

Subject #57 gave 13 definite responses; 4 of them were in accord with an explicit link in the problem, and the remaining 9 classified by shortest path reasoning. Similarly, subject #18 gave 12 definite replies, 4 of them agreeing with explicit link acceptance, another 3 consistent with shortest path reasoning. Two of this subject's responses were in accord with the polarity of the only path through the network under the inclusion assumption regarding negativity, and 3 additional definite replies could not be classified by any reasoning strategy we have so far considered. Finally, subject #54 gave 3 answers agreeing with explicit link acceptance, 4 agreeing with shortest path reasoning, 4 agreeing with most paths reasoning, 3 which were in accord with the polarity of the only path given the inclusion assumptions, and 1 more which was not predicted by any strategy we've addressed.

Overall among the subjects who consistently acted in agreement with the literature's assumption that negative chains are indeterminate shortest path reasoning (with explicit link acceptance as a special case) seemed to be the most frequent predictor of subject response. Assuming that the classification of negative chains also applied in larger graphs, these subjects tended to agree with transitivity applied to the remaining links. Most path reasoning was a reasonable predictor where shortest path reasoning was not. While this demonstrates that the reasoning patterns identified under the stringent reasoning criteria are also apparent elsewhere in the data where the transitivity requirement is relaxed, this relaxation still generates only a very small pool of subjects to consider.

2.2.2 Relaxing Negativity

Holding transitivity fixed but allowing increasing amounts of determinate responses to negative chains yields 20 subjects. Table 3 encapsulates the between subject patterns of reply for those cases where significance obtains.

A within-subject analysis reveals that there are 9 skeptics. Of these, 5 were already discussed in § 2.1.1, before the relaxation of the inclusion criteria, as well as the affirmer, and the explicit link acceptor. There were 3 whose behavior was consistent with preferring shortest path reasoning in cases in which it could resolve the problem, and most paths reasoning otherwise (one of these was #12, also discussed above). An additional 3 were best predicted by the reverse preference: most paths reasoning and then shortest paths. A final 3 behaved as if guided by most paths reasoning (and are not predicted by the acceptance of explicit links). The

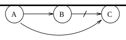
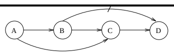
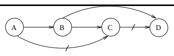
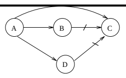
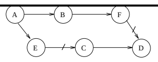
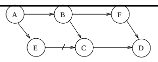
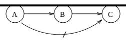
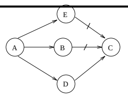
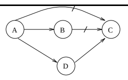
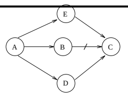
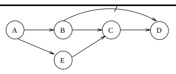
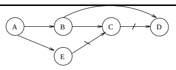
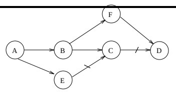
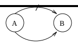
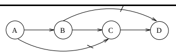
Problem	Response		
	a	b	c/d
2 	6	1	14
7 	6	0	15
8 	7	3	11
9 	6	1	14
16 	0	17	4
17 	11	0	10
18 	0	7	14
22 	1	5	15
26 	1	8	12
28 	4	1	16
31 	6	2	13
33 	5	0	16
35 	4	0	17
38 	0	0	21
40 	1	5	15

Table 3: Negativity Relaxed, Between-Subj. Patterns

average number of definite answers for these subjects is 6.86 ($\sigma = 6.07$). Thus, apart from affirmation, we see in this relaxation of the inclusion criteria the same patterns of response that were found among the smaller set of absolutely consistent subjects. Increasing use of most paths reasoning accompanying increased likelihood to classify negative chains negatively.

2.2.3 Degrees of Transitivity and Negativity

For the final set of analyses, we considered simultaneous relaxation of both the transitivity and the negativity requirements. We examine the results according to participation of subjects in groups defined by a certain degree of consistency. We with the exception of affirmation, we find the same representation of reasoning strategies in each group, although with less consistent internal applicability as the groups tend towards greater deviation from the inclusion criteria. The groups, as a function of degree of meeting the inclusion criteria, broke down into 5 classes, encompassing 51 subjects eligible for consideration through being systematic in their responses. The inclusion function was defined from a matrix with increasing intransitivity as one axis and decreasing negative chain indeterminacy as the other axis. The group that a subject falls into is then determined from this matrix.[8] Below we consider variances among these groups.

For these purposes, it made sense to make the strict criterion even more strict, so that the one subject considered above (#12) who gave a positive definite classification to a negative chain with 2 negative links was instead placed in the second group.[9] This means that Group 1 is comprised of 4 complete skeptics, one explicit link acceptor and the one affirmer. Also, this entails that the revised Group 1 is the only group comprised solely of subjects in just the graphical conditions. In the remaining groups there was roughly an even distribution between subjects in the sentence only condition and subjects in the graphical conditions. An exception is Group 2 which had 5 subjects in the sentence-only condition. That so many subjects from this (the most abstract and taxing) condition were in the group second-most consistent with respect to foundational assumptions of inheritance is a concrete support for the rationality of those assumptions. This section proceeds by summarizing the results of an analysis like the detailed one given in above for the Group 1 subjects—we examined the within-subject patterns of response, but report just the between-subject averages of those within-subject patterns.

Group 2. The 6 subjects in this group tended to respond definitely to more problems than subjects in Group 1 (there the average was 5.9 (σ=8.2), apart from the *affirmer* who answered 23 definitely, rather exceptionally for the study as a whole); in Group 2, the average was 10.6 (σ=4.8) definite classifications per subject, beyond those required by the inclusion crite-

[8] If i is the group number, then subject n is in group i if that subject's number of 'incorrect' responses to transitivity and negative chain test cases falls into the region of the matrix defined by the $g(i)$ cells of the matrix closest to the origin uniquely corresponding group i. This function is defined as follows: $g(0) = f(0) = 0$, $f(i) = i^2 + i$, $g(i) = f(i) - f(i-1)$; essentially, this describes a regular series of encompassing rectangles. There are 2 cells in the first group, one empty as there were no subjects with all 7 negative chain problems 'correct' but one of the transitivity problems 'wrong'.

[9] The subjects considered in the second absolute criterion (Vogel, 1996), do not appear in the following analysis. This is because those subjects were required to rate *all* of the negative chain criterion problems contrary to the literature's predictions. According to the present criterion, those subjects all fall into group 8, and are thus outwith discussion. On the other hand, subjects from the first absolute and its relaxation (in which negativity was held as the literature predicts and transitivity was relaxed) do participate here in groups 2 and 5.

rion, with no outliers.[10] These subjects' responses are classifiable in terms of the same basic reasoning strategies for those problems classified definitely as described for Group 1. Two of the subjects' definite responses were mainly those predicted by shortest path reasoning (8 problems) and the remainder by explicit link acceptance (a flavor of shortest path reasoning; 3 and 4 problems, respectively). One of these 2 subjects is #12, whose responses were detailed above. Moreover, there was a high rate of consistency between these subjects with respect to the problems they found classifiable in the way they did. Of the 4 remaining subjects in this group there was a more even distribution of correspondences to the three strategies. One of these subjects gave positive definite classification to 4 additional problems from which a definite reasoning pattern cannot be abstracted.

Group 3. There were 6 subjects in this group as well, with an average of only 5.2 definite answers (σ=3.3) (one subject was deemed an outlier in this group and excluded from further analysis, including the just-mentioned average: this subject did not give answers to 16 of the problems; 6 of those were among the inclusion criteria (which include this subject because inappropriate answers were not given) and 10 among the remaining 25 problems). One subject's (#75) definite responses were all predicted by explicit link acceptance. The remainder were split between explicit link acceptance, shortest path reasoning proper, most paths reasoning, and non-predicted reasoning. One of the latter 2 subjects was in the sentence-only condition and gave 2 answers which upon inspection were consistent with the polarity of the first path traceable as a chain of sentences, given the order presented.

Group 4. This group marks a transition into greater inconsistency. It contains 16 subjects, with an average of 11 problems each (σ=5.0) classified definitely, beyond the inclusion criteria. Among those, most paths reasoning was most frequently the best classification of responses (on average, 2.3 times per subject (σ=1.3)). Explicit link acceptance was the best predictor 1.1 times per subject (σ=0.8), and shortest path reasoning proper, 2.0 times (σ=1.3). Of the remaining problems classified definitely, an average of 2.9 (σ=2.4) could not be easily predicted by any current theory. There was an even distribution of subjects in this group between graphical and sentential conditions. The subjects with many problems answered 'strangely' definite tended to be in the sentence-only condition. Their responses, on inspection, coincided with the polarity first chain of sentences available, according to the order in which they were presented. In sum, this group also tended to skepticism, and most path reasoning was the best classification of definite responses.

Group 5. Although the average number of problems per subject given definite responses was 9.7 (σ=5.3), Three skeptics made 3 or fewer definite classifications (two were in the graph-only condition, and one in the sentence-only condition, with 10 of the 17 total subjects in this group in sentence only, and 7 in the graphical conditions). Of the remaining subjects, 1.2 (σ=1.3) was the average number of times/subject that explicit link acceptance was predictive; 2.2 (σ=1.6), shortest path reasoning, 1.8 (σ=1.2) most paths reasoning. There were 3.1 problems on average (σ=2.3), per subject which were not classifiable by any known theory of inheritance. Again, these tended to follow from subjects participating in the sentence-only condition, and were consistent with the polarity of the first path constructible between the queried classes, according to the order in which the sentences were listed. Again, in sum, we have support mainly for skepticism, and among definite (nonarbitrary) responses, for a combination of shortest path and most paths reasoning. Note that group 5 is the closest group within consideration to the subjects analyzed in § 2.1.2 in terms of relaxing the negativity criterion. That is, subjects in group 5 are much more likely than group 2 (say) to rate negative chains as negative. Examining the answers of subjects in this group in light of the possibility that they tended to consistently rate negative chains as negative yields 3.4 (σ=2.2) as the average number of times most paths reasoning was applicable.

2.2.4 Analysis

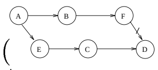
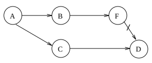
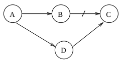
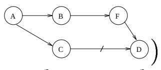
Now we examine the commonalities and variances among the groups. There were 8 groups altogether, but, as mentioned, after the level of relaxation represented by group 5 we did not find significant consistency. Taken together, groups 1, 2 and 3 had roughly the same number of subjects as each of groups 4 and 5 individually had; thus, we examine patterns which obtain among the regrouping g123, g4 and g5.[11] However, we also found roughly similar degrees of skepticism in groups 1, 2 and 3, and a significantly different (Tukey,[12] $p < .05$) degree of similarity between 4 and 5, so for other purposes we will also compare g123 with g45. Both of the regroupings are in simple supersets of the initial groupings described above.

First we mention systematicities across all groups. For example, there was not a significant difference in efficacy of explicit link acceptance in any of the groups. Also, there is not a significant difference in indeterminate ratings of 'classically' ambiguous problems

[10] Recall, such figures refer to problems other than those which defined inclusion.

[11] We hope the naming scheme is obvious — g4 is just group 4; g123 is the sum of groups 1, 2 and 3; and so on.

[12] This is the Tukey H. S. D. test, useful for *post hoc* analysis of differences in means; it is a more critical test than a t-test, for instance. Henceforth, in this section the name will be dropped and only the corresponding measure of significance given.

(, ,  , ). As an interesting example, among the 51 subjects who comprised the 5 groups, 77.6% rated the Nixon diamond indeterminate, 6.12% gave it a positive definite classification, and over double that, 16.3%, gave the problem a definite negative classification.[13] Both of these constancies serve to emphasize the generality of the consistency patterns observed in the original very small sets of absolutely consistent reasoners.

However, there remains evidence suggestive of partitions in reasoning strategies corresponding to the groupings that we defined by degrees of satisfaction of the inclusion criteria. First note that while a significant increase in 'arbitrary' responses does not occur through all 5 groups, it does obtain between g123 and both g4 ($p < .05$) and g5 ($p < .02$), as well as between g123 and g45 ($p < .01$). Also, there is a significant trend towards determinate responses between g123 and g45 ($p < .05$). Both of these facts should be anticipated by the very nature of the group definitions. For shortest path reasoning there is a very high mean of within subject applicability of the strategy for subjects in group 2. If explicit link acceptance is included as accepting the shortest path, then the difference in means for group 2 and each of the other 4 is significant (g1: $p < .05$; g3, $p < .01$; g4, $p < .05$; g5, $p < .05$). Group 2 seems to be best described then as shortest path reasoners. As for most paths reasoning, there is a significant difference in means between groups123 and both group4 ($p < .01$) and group5 ($p < .05$), as well as one between the combined g123 vs. g45 ($p < .01$), with a rise for the number of problems that could be classified by most paths reasoning for the more inconsistent groups. As the negative reasoners in Vogel (1996) also could be classified best as most paths reasoners this could point at a correlation between increasing negative chains rated negative and most paths reasoning.

The attentive reader will recall that the answers predicted by the various reasoning strategies agree in some cases with predictions made by systems that exist in the literature. Indeed, roughly the same patterns of between-group significance obtain for H90 as does for the shortest-path reasoner we've described (e.g. distinguishing group 2 from the others, and setting g123 as more in accord with the predictor than g4 or g5). However, these results follow almost exactly from the problems on which H90 agrees with shortest path reasoning. In fact, the shortest path reasoner was a strictly better predictor (by 50%), on the problems for which the two reasoners predicted different responses. The skeptical reasoner, different from H90 in allowing ambiguities to cascade and in demanding only on-path preemption, does in fact predict answers quite well for groups 1, 2 and 3, in comparison to H90, mainly because it is skeptical. We've already seen that subjects in those groups are more skeptical than the others. We've also claimed that the definite classifications of group 2 in particular is best predicted by shortest path reasoning. The cases in which skeptical reasoning out performed shortest path reasoning were also just those which generated indeterminate classifications.

[13] In comparison, the pilot study reported by Elio and Pelletier (1993) had 50% indeterminate response to the Nixon diamond, with the remainder split equally between positive and negative definite classifications. The between-subject analysis in the initial experiment of Hewson and Vogel (1994) yielded: 65% ind., 15% pos., 19% neg.

3 FINAL REMARKS

We have identified five reasoning strategies based on patterns of response within individual subjects classifications of the problems. Subjects responses were mainly skeptical, skeptical but accepting of explicit links, accepting of shortest paths generally, and/or accepting of most paths reasoning. The result is a family of reasoners which can, on the basis of evidence so far collected, be regarded as psychologically plausible. We have given details of the structure available in the data which supports these judgements. We found that the basically consistent subjects were roughly split between being very skeptical and less so (having found in the context of these experiments that people behaved mainly in accord with skepticism). Among the more skeptical, shortest path reasoning was the best predictor of definite responses. Among the less skeptical, and particularly among those who rate negative chains as definitely negative rather than as indeterminate as the literature expects, most paths reasoning was the best fitting predictor.

We emphasize that this has been an exercise in mining the data accumulated by Hewson and Vogel (1994) and Vogel (1995). While those experiments were designed with in mind the sorts of scrutiny we have engaged in here, we also investigated other issues which were apparent in the data by surprise. This constitutes a caveat with respect to accepted psychological methodology, and certainly future experiments we conduct will be obliged to investigate these issues directly from the outset. Nonetheless, the clarity of the results and the degree to which these conclusions are supported let us feel secure that they are not actually based on spurious significances.

This paper has reported within-subject analyses of the data accumulated during the initial experiments investigating the plausibility of inheritance reasoning as a model of human reasoning with defaults. Our ongoing work in this area includes replications and a battery of other experiments to clarify some of the issues that have been opened during analyses like the present one. For example, we are very interested in obtaining a clearer understanding of the role of mode of presentation. The current results hint at a more polarized response among those subjects who received the problems in a graphical condition than those who had only the sentential presentations of problems to work with, however, we do not yet understand how to interpret these facts, particularly since the same basic patterns of response that we mention here obtained in both the sentence-only condition and the graphical conditions. The exception is that group 2, which was mainly comprised of subjects in the sentence only condition, was the group that also best satisfied the predictions of shortest path reasoning. We lack evidence suggesting that these two facts are related, but it is a fact which begs further investigation.

Additionally, we are designing another style of experiment which involves a task following the classification of a problem and depending on the classification made. We hope that in such a context we can induce a greater proportion of definite classifications, since there will be a need to act on the conclusions which was lacking in the experiments discussed here. The differences in patterns of response, if any, will be rather interesting to learn. One of the problems with the design of the experiments which generated the data analyzed here is that the questionnaire was rather long, and the time required to deal with it seriously would certainly have encouraged frequent use of the psychologically 'easiest' of the available answers: indeterminate. Of course, we cannot enumerate here the full range of experimental questions that this work has raised in our minds. We hope, though, to have made clear that there is an extremely interesting set of issues to consider through further experimentation, and that the result will be informative to researchers who desire to build psychologically plausible formal models of default reasoning. We have offered preliminary suggestions for nonmonotonic reasoning strategies that people might be using in some circumstances, but our basic point is that plausibility judgements should not derive solely from a logician's introspection on examples but rather from observation of consistent rational behaviors that people actually exhibit.

Acknowledgements

Deep thanks to Robin Cooper, Claire Hewson, Jon Oberlander and Jeff Pelletier. Vogel is grateful for the Marshall Scholarship that let him study at the Centre for Cognitive Science in Edinburgh, and to the German SFB 340 for allowing him to work in Stuttgart.

References

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.

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.

Evans, J. S. B. T. (1983). Linguistic Determinants of Bias in Conditional Reasoning. *Quarterly Journal of Experimental Psychology*, *35*(A), 635–644.

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

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., 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.

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.

Rayner, K. & Pollatsek, A. (1989). *The Psychology of Reading.* Englewood Cliffs, NJ: Prentice Hall.

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.

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.

Vogel, C. M. (1995). *Inheritance Reasoning: Psychological Plausibility, Proof Theory and Semantics.* Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Vogel, C. M. (1996). Human Reasoning with Negative Defaults. In Gabbay, D. & Ohlbach, H. J. (Eds.), *Practical Reasoning*, pp. 606–621. Lecture Notes in Artificial Intelligence 1085. Berlin: Springer Verlag. Proceedings of the *International Conference on Formal and Applied Practical Reasoning, FAPR'96.* Bonn, Germanny, June 1996.

Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content.* Cambridge, Mass.: Harvard University Press.