On Reasoning with Default Rules and Exceptions

Renée Elio
Department of Computing Science
University of Alberta
Edmonton, Alberta T6G 2H1
ree@cs.ualberta.ca

Francis Jeffry Pelletier
Departments of Philosophy and Computing Science
University of Alberta
Edmonton, Alberta T6G 2H1
jeffp@cs.ualberta.ca

Abstract

We report empirical results on factors that influence how people reason with default rules of the form "Most x's have property P", in scenarios that specify information about exceptions to these rules and in scenarios that specify default-rule inheritance. These factors include (a) whether the individual, to which the default rule might apply, is similar to a known exception, when that similarity may explain why the exception did not follow the default, and (b) whether the problem involves classes of naturally occurring kinds or classes of artifacts. We consider how these findings might be integrated into formal approaches to default reasoning and also consider the relation of this sort of qualitative default reasoning to statistical reasoning.

Introduction

Default reasoning occurs whenever the evidence available to the reasoner does not guarantee the truth of the conclusion being drawn; that is, does not deductively *force* the reasoner to draw the conclusion under consideration. For example, from the statements 'Most linguists speak more than three languages' and 'Kim is a linguist', one might draw the conclusion, by default, 'Kim speaks more than three languages'. Subsequent information may force the reasoner to withdraw that conclusion; default reasoning is also termed non-monotonic, because the sentences held true at time 1 may not be true at time 2. We will call "Most linguists speak more than three languages" a default rule.

If an artificial agent were to wait for the information necessary to draw an inference sanctioned by classical deductive logic, then no conclusion might ever be drawn. Much of what is considered to true in the world is true only most of the time: there are exceptions and sometimes interacting default assumptions that can lead to conflicting conclusions. A good deal of work has been done in the AI community at formalizing default reasoning, either through qualitative approaches using conditional logics (e.g., Delgrande, 1987), probabilistic approaches (e.g., Bacchus, 1991), or approaches that attempt to capture quantitative notions within a qualitative framework (Gefner, 1992; Pearl, 1989). In the last several years, there has been an increasing attention in the default-reasoning community given to formalizing the notions of relevance and irrelevance, i.e., what information would be (ir)relevant to deciding whether a default rule applies in a particular case (see Greiner & Subramanian, 1994). For example, these frameworks propose ways of assessing the (ir)relevance of Kim's membership in the class of "red-haired people" to the application of the three-languages default rule and similarly, of Kim's membership in the class of "graduates of University X"—about which there may be a conflicting default rule about language skills. In the latter case, default reasoning theories aim to identify general and consistent means of specifying which of possibly several conflicting default rules should apply to an individual.

Generally speaking, the knowledge of *other* exceptions to a default rule has not yet been a factor in whether a particular default rule applies in a given case. As we see below, information about known exceptions to a default rule are not "supposed to" influence the application of that rule in a particular case. The studies we report are a continuation of previous work (Elio & Pelletier, 1993) aimed at understanding how people reason with rules that have exceptions, and what factors influence people's application of those rules. How people reason with default rules and exceptions *per se* has not received much attention within the cognitive psychology community (see, however, Collins & Michalski, 1989). However, there are overlaps between the issues we investigate in this work and those that have been considered in the literatures on statistical and inductive reasoning by people. We highlight some of the relationships we see in the sections that follow.

Benchmark Problems on Default Reasoning

Table 1 presents the a subset of the problem types that we used in this study. These problems were taken from the so-called "Nonmonotonic Benchmark Problems" (Lifschitz, 1989). These benchmarks formalized types of non-monotonic reasoning and specified the answers generally accepted by AI researchers in the area and which any non-monotonic theory was supposed to validate. Put another way, these are the defined "correct answers" for problems that take this form, despite some acknowledged difficulties in deciding just what the correct answers should be (Touretsky, Horty, & Thomason, 1987). Elsewhere, we have argued that, unlike human performance on symbolic deductive logic problems, the kinds of default conclusions people draw actually defines phenomenon of interest to be achieved by artificial agents; and thus empirical data on

1	Blocks X and Y are heavy. Heavy blocks are normally on the table. X is not on the table		2	Blocks X and Y are heavy. Heavy blocks are normally on the table. X is not on the table. Y is red.	
	Q: Where is Block Y?	*A*: on table		*Q*: Where is Block Y?	*A*: on table
3	Blocks X and Y are heavy. Heavy blocks are normally on the table. Most heavy blocks are red. X is not on the table. Y is not red.		4	Blocks X and Y are heavy. Heavy blocks are normally on the table. Block X might be an exception to this rule.	
	Q: What color is Block X? *Q*: Where is Block Y?	*A*: red *A*: on table		*Q*: Where is Block Y?	*A*: on table

Table 1: Four default reasoning problems and benchmark answers

human default reasoning has an important role to play in validating default reasoning theories and in identifying principles by which default answers can be assessed (Pelletier & Elio, 1995). It is this rationale that motivates our interest in understanding factors that influence people's default conclusions on even these simple problems.

We call the four problems in Table 1 the "basic default reasoning problems." They concern two objects governed by one or more default rules. Additional information is given to indicate that one of the objects (at least) does not follow one of the default rules. We refer to this as the *exception object* (for that default rule) or *default violator*. The problems then ask for a conclusion about the remaining object. We refer to this as the *object-in-question*. It is apparent from the sanctioned benchmark answers for these problems that the existence of known default violator, or any additional information about the object-in-question (e.g., Problem 2 in Table 1), should have no bearing on a conclusion drawn about the object-in-question when using that rule.

Experiments on Basic Default Reasoning Problems

In previous studies on these sorts of problems (Elio & Pelletier, 1993), we reported evidence suggesting that people's plausible conclusions about defaults and exceptions are influenced by the apparent similarity between a given default violator and the object-in-question. We were naturally lead to wonder just what kind of similarity mattered to deciding whether or not some object follows a default rule or instead behaves like a known exception. Our conjecture was that the similarity to a default violator may be relevant when the shared features could account for why the exception object violated the default rule in the first place. If the object-in-question also has those features, then it too may behave like the known exception and also violate the default rule. The results we report below are further investigations of those findings.

Design

We defined three conditions in which to present the four canonical default reasoning problems given in Table 1: (a) a no-shared features condition, (b) a superficial shared-features condition, and (c) an explanatory shared-features condition. In the superficial case, the objects were described as having certain features in common; these features corresponded to those given by subjects in a separate norming study as irrelevant to the conclusion offered by a default rule. Typically, these were physical features for the actual cover stories (example below) that we used for the problems. The explanatory shared-features corresponded to features given by subjects in the norming study as relevant to the conclusion implicated by a default rule; these explanatory features typically concerned an object's use or function. The hypothesis was that subjects would apply the default rule to the object-in-question most often when there was no information about its similarity to the default violator, and least often when the common features between the object-in-question and the default violator could support an explanation of why the default-violator itself did not obey the default rule. The superficial condition should lie somewhere in-between.

Figure 1 illustrates this manipulation for Problem 1. For all problems, the order of information was: the set-up sentences, marked (a) in Figure 1; the sentences corresponding to the similarity information (if any), which are marked (b') and (b") for the two similarity manipulations; the default rule, marked (c); the sentence marked (d) indicating the rule violator did not follow the default rule; and finally the question (e) asking for a plausible conclusion about the object-in-question. In addition to the medical journals scenario, there were cover stories about membership in university clubs, distribution of student ID cards, and operations of campus parking lots. Similarity was a between-subjects factor and problem type was a within-subjects factor. Subjects saw each of the four benchmark problems under one type of similarity, with each benchmark having one of the four possible cover stories. The assignment of cover-story to each problem type was counterbalanced across subjects.

Subjects and Procedure

Seventy-two subjects were randomly assigned to one of the three similarity conditions. The problems were randomly

No Similarity

(a) Cardiac News and Drug Developments are medical journals you need for a research paper.

(c) Medical journals are usually located in the Health Sciences library.

(d) Cardiac News is an exception: It is not in the Health Sciences library—It is kept in the Department of Medicine Reading Room.

Superficial Similarity Additions

(b') Both Cardiac News and Drug Developments are published in Canada. New issues of both journals come out every month. They are bound in light-blue covers.

Explanation Similarity Additions

(b") Both Cardiac News and Drug Developments are among the most expensive journals the university purchases. There have been problems with stolen or missing copies of these journals over the years. Both of them are consulted on a daily basis by graduate students in Medicine.

Question

(e) What would be reasonable to conclude about where Drug Developments is located?

Figure 1: Components of alternative similarity versions for problem type 1

ordered in booklet form. Each problem's question (see Figure 1) was followed by four possible answers, corresponding to these options (tailored to each cover story): (a) the object-in-question followed the default rule, (b) the object-in-question violated the default rule, (c) no conclusion was possible (a "can't tell" option), and (d) "other", for which subjects could write in another conclusion. The instructions emphasized that we were interested in common-sense conclusions, and that there were no right or wrong answers.

Results

The data from three of the 72 subjects had to be discarded, due to a mis-assignment of experimental materials. This left a total of 69 subjects, 23 in each of the three similarity conditions. Table 2 shows the proportions of each answer category as a function of answer category and similarity level.

Because the data we collected are interval data, i.e., answers falling into one of four response categories, they do not necessarily follow a normal distribution. One appropriate treatment of such data is a loglinear analysis of models defined by particular combinations of main effect and interaction terms. Under this approach, we evaluate whether a given model's predicted data is significantly different from the observed data, using a χ^2 likelihood ratio statistic. A model with fewer terms (and more degrees of freedom) is preferred to a model with more terms, provided that the predicted data does not differ significantly from the observed data. The simplest model we identified included a main-effect term for answer category and an answer-category by similarity interaction term ($\chi^2 = 32.48$, df=38 , p = .722). If the interaction term is removed, the difference between observed and predicted data approaches significance ($\chi^2 = 72.53$, df=56, p = .068).

It is clear from Table 2 that, most of the time, subjects applied the default rule to the object in question (the model's main-effect term for answer category) and it is also apparent that this decision was influenced by the apparent similarity to another object that violated the rule (the model's interaction term). The trend in the frequencies of applying the default rule to the object-in-question was in line with our predictions, occurring least often in the explanatory condition. We note that subjects were conservative in their reluctance to apply the default rule in this case, choosing the "can't tell" (.21) option rather than the explicit rule-violation option. We cannot account for the tendency for subjects in the superficial condition to provide so many "other" conclusions. Although the superficial features were identified from a norming study as being irrelevant to the property implicated in the default rule, it is possible they were not. Hence, a possibility remains that subjects tended to reject the default rule given any information they could use to construct an alternative prediction about the object-in-question. A laboratory manipulation of inter-object similarity may be weaker than tapping into extant knowledge of similarity between object classes; this is a line of investigation we are currently following. Still, these results are consistent with our previous findings that the application of a default rule may be influenced by information about other exceptions to the rule.

		Answer Category			
		Follows Default	Violates Default	Other	Can't Tell
Similarity:	none	.70	.12	.11	.07
	superficial	.54	.18	.25	.03
	causal	.45	.19	.15	.21

Table 2: Proportion of Responses as a Function of Similarity and Response Type

Birds-Fly Context		Birds-&-Bats Fly Context	
Animals normally do not fly		Animals normally do not fly.	
Birds are animals. Birds normally fly.		Birds are animals. Birds normally fly.	
Ostriches are birds.		Bats are animals. Bats normally fly.	
Ostriches do not fly.		Ostriches are birds. Ostriches do not fly.	
Q: Do birds other than ostriches fly?	*A*: Yes	*Q*: Do birds other than ostriches fly?	*A*: Yes
Q: Do animals other than birds fly?	*A*: No	*Q*: Do animals other than birds & bats fly?	*A*: No

Table 3: Two default inheritance problems

Reasoning about Inherited Default Properties

In Table 3, we present two additional problems from Lifschitz's (1989) nonmonotonic benchmark set. These problems are easily recognized as canonical examples of conflicting default knowledge about classes related in a class-subclass hierarchy. These problems were included in Lifschitz's benchmark set because they capture several essential questions that have been central to reasoning theories about classes, subclasses, and individuals, namely how should properties—some of which are definitional and some of which are prototypical— be "inherited" by the next element down the hierarchy? Other more complex inheritance scenarios are accommodated by different formal default reasoning theories, but these problems present simple cases of conflicting default rules.

In some previous pilot work, we found that subjects generally allowed the default properties to be inherited, as per the "correct" answers given in Table 3. In this study, we examined whether this application of default properties was sensitive to the kind of taxonomic categories being considered, namely *natural kind categories* or *artifact categories*. The notion that "kinds" influences reasoning has been considered in both the inductive inference and the statistical reasoning literatures (Thagard & Nisbett, 1993). People's tendency to reason statistically can also be influenced by perceived variability and homogeneity in the classes they are considering. For example, Nisbett et al. (1983) reports that people expect a lower variability for natural classes than for classes of human behaviors. Hence, it seemed to us that this kind of metaknowledge, implicated in some statistical reasoning studies, may also impact upon qualitative judgments concerning the inheritance of default properties.

The second factor we manipulated was whether the problems included class-size information for the classes and subclasses that formed the inheritance hierarchy. Our inclusion of this factor was also motivated by our desire to bridge these qualitative default reasoning decisions with some statistical reasoning results, that have indicated that people are influenced by class size information in making some kinds of inferences (Nisbett et al., 1983). For this initial study, we contrasted a *class size absent* case, in which there was no mention of how large the subclasses were, with a *class size present* case. In this latter condition, the problem mentioned particular figures for class sizes, falling within the 20-80 range. Our contrast of these two cases here was not to assess how particular class-size values would lead to different conclusions; rather, we wanted first to assess whether framing these qualitative inheritance problems in a somewhat more quantitative guise would influence people's tendency to ascribe the inheritable default properties to particular subclasses.

Design

Each subject received each of the two inheritance problems with both a natural kinds cover story and an artifact cover story. For the natural categories, we used stories about trees and snakes; the artifact categories concerned taxation laws for cigarettes and features of medieval musical instruments. No cover story was repeated in any of the problems that a subject saw, and the assignment of cover stories to problems was counterbalanced across subjects. In addition to these four problems, subjects solved two other inheritance problems that were part of a different study.

Subjects and Procedure

Sixty-four subjects were randomly assigned to receive either the class-size present problems or the class-size absent problems. The problems were presented to subjects as "short paragraphs that were extracted and adapted from newspapers and popular science articles [which presented] some facts but left other information unstated." Subjects were told that their task was to specify the reasonable, common-sense conclusion or inference they would draw, based strictly on the information given to the reader in the excerpts presented. Below is the text of a birds-fly context problem, using natural categories and including class size information:

> ....The kind of trees you plant can also help attract birds year round. Coniferous trees do well in our region. Unfortunately, most of the 63 species of coniferous trees produce a bitter sap. An example is the subclass "cedrus" (cedar): taste the sap from any type of cedar tree and you will be unpleasantly surprised at how bitter it is....
>
> There is, however, a subclass of coniferous trees called "Pinaecea" that any good garden nursery will know about. Most of the 22 Pinaecea species give sweet sap that attract squirrels and certain types of birds.... [however] one Pinaecea tree to be avoided is *picea mariana*: it gives bitter sap. An attractive and

		Class Size Present	Class Size Absent	Mean
		Natural Classes		
Do birds other than ostriches fly?(yes)	birds-fly context	.72	.72	.72
	birds-&-bats-fly context	.69	.69	.69
Do animals other than birds fly?(no)	birds-fly context	.69	.71	.71
	birds-&-bats-fly context	.59	.56	.58
		Artifact Classes		
Do birds other than ostriches fly?(yes)	birds-fly context	.47	.69	.58
	birds-&-bats-fly context	.41	.69	.55
Do animals other than birds fly?(no)	birds-fly context	.59	.63	.61
	birds-&-bats-fly context	.63	.72	.68

Table 4: Proportion of "Correct" Answers Given for Inheritance Problems from Table 3

> hardy pine tree, make sure your local garden nursery doesn't try to sell you this one if youraim is to attract local wildlife....

The class-size absent versions of these problems replaced references to statements like "Most of the 63 species" with "Most species" Two questions then followed about subclasses that the alleged article did not mention. For the above example, these questions were:

> z(a) "The article mentioned one subclass of conifers—Pinaecea —but did not discuss a second subclass called Juniperous. From the information presented in the article, what is your common-sense conclusion about whether species within the Juniperous subclass produces sweet sap or bitter sap?" [*do animals other than birds fly*?]
>
> (b) "The article failed to mention another Pinaecea tree called *libani chrysolitis*. From the information presented in the article, what is your reasonable conclusion about the kind of sap it produces?" *[do birds other than ostriches fly?]*

Subjects selected their answer to each question from one of three possibilities that corresponded, in essence, to "yes [it can fly]", "no [it can't fly]" and "can't tell."

Results

Even though subjects did not solve problems mentioning birds, bats, and flying, it is easiest to talk about the results by referring to these canonical terms. Table 4 presents the proportion of subjects choosing the prescribed default answer for each of the two possible questions. A loglinear analysis of the data identified a model defined by three interaction terms: category type X quantifier X question X answer category; category type X context type X answer category, and context type X question X answer category ($\chi^2 = 2.89$, df = 15, p = 1.0). We cannot at this stage propose an account for all these interactions, particularly those involving the birds-fly vs. birds-plus-bats fly context effect. However, the response patterns that give rise to this model are evident in Table 4. First, the proportion of prescribed default-inheritance answers was higher when subjects were reasoning about natural categories and lower when they were reasoning about artificial categories. Second, subjects were less likely to allow the default property to be inherited for artifact classes than for natural classes, particularly when the problems mentioned class size information. This finding is consistent with the notion that people may perceive artifact/artificial classes as inherently more variable than natural kinds and that this impacts their willingness to ascribe default properties. The impact of merely mentioning class sizes may have triggered this consideration of variability. Unfortunately, this sort of conjecture does not seem entirely consistent with the effect of the birds-fly versus birds-&-bats-fly context for the natural classes condition. For those problems, subjects were less likely to apply the default rule to conclude that animals-other-than birds don't fly, when bats were included as a second exception. This did not occur with the non-natural stimuli; whether the number of known (or salient) exceptions influences the application of a default rule needs further study.

Discussion

These results suggest an aspect of plausible reasoning that is missing from current non-monotonic theories, namely that there are certain kinds of information that are relevant to applying default rules. The findings outlined above suggest some considerations about what is relevant in non-

monotonic reasoning: inter-object similarity, natural vs. artificial categories, and class-size information. If an unknown case is similar to an understood exception, then a plausible conclusion may not be to apply the default rule, but instead to predict that the unknown case behaves instead like the exception. The influence of the latter two factors border on a meta-knowledge effect: people may work certain assumptions about class variability into their default conclusions, e.g., there is greater regularity to how defaults and exceptions operate in the "natural" world than there is for non-natural classes and subclasses. Other researchers have appealed to this distinction in the realm of inductive reasoning (e.g., Thagard & Nisbett, 1993) and statistical reasoning (Nisbett et al., 1983). Our results in this study are suggestive, though by no means conclusive, that this distinction may come into play in the sort of qualitative default reasoning scenarios we have studied here. Indeed, another interpretation from the inheritance problems is that, in some arenas, human reasoners are much more cautious in their attribution of default properties than we might otherwise believe.

How can our findings, that knowledge about known default-rule exceptions may influence the application of default rules to similar cases, be worked into a specification for non-monotonic reasoning? One idea might be called "explanation-based default reasoning", in the same sense of this notion used in the machine learning literature. That is, a reasoner attempts to explain why some default rule does not apply to the known exception, and then evaluates whether that explanation applies to the object under consideration. This emphasis on reasoning about one known individual case in order to make a decision about another case may sound like there can be no formalized rules for default reasoning. But this account need not be taken as a prescription for a strictly case-based approach to default reasoning. First, the influence of a similar object might be used to direct the selection of an appropriate "reference class," about which some statistical properties could be inherited. Second, some formal theories already appeal to the notions of causality, explanation, or argumentation processes to construct, and then select among, alternative models that are defined by conflicting default rules (e.g., Pollock, 1987; Gefner, 1992). Along these lines, we note that the presence of a known rule violator in the problems we investigated here may be a red herring, insofar as the important aspect for subjects may have been the availability of information that could support an explanation about why a default rule may not apply (quite independent of whether some other known violator was salient). In general, it does not seem plausible that additional information about an individual should necessarily have no impact on determining whether it follows a default rule (e.g., whether a block's color influences whether it is, by default, located on a table). One interpretation of our findings is that such information is relevant to the extent it supports explanations for (or against) a default assumption; a known exception to a default may serve as a touchstone for constructing these explanations.

Acknowledgments

This work was supported by Government of Canada NSERC Research Grants A0089 (RE) and A5525 (FJP). We thank Siobhan Neary for her assistance in conducting these experiments and the University of Alberta Department of Psychology, for their continuing cooperation in allowing us access to their subject pool.

References

Bacchus, F. (1991). *Representing and Reasoning with Probabilistic Knowledge*. Cambridge: The MIT Press.

Collins, A. & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science, 13*, 1-49.

Delgrande, J. (1987). An approach to default reasoning based on a first order conditional logic. In *Proceedings of AAAI-87*, 340-345, Seattle.

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). Boulder, CO.

Gefner, H. (1992). *Default reasoning: causal and conditional theories*. Cambridge: MIT Press.

Greiner, R. & Subramanian, D. (1994). *Relevance: American Association for Artificial Intelligence 1994 Fall Symposium Series*, November 4-6, New Orleans. Palo Alto: AAAI Press.

Lifschitz, V. (1989). Benchmark problems for formal nonmonotonic reasoning, v. 2.00. In M. Reinfrank, J. de Kleer, & M. Ginsberg (Eds.) *Nonmonotonic Reasoning* 202-219, Berlin: Springer-Verlag.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday reasoning. *Psychological Review, 90*, 339-363.

Pearl, J. (1989). Probabilistic semantics for nonmonotonic reasoning. *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, 505-516, Toronto, Canada.

Pelletier, F. J. & Elio, R. What should default reasoning be, by default? (Tech. Rep. 94-13) Edmonton: University of Alberta, Department of Computing Science.

Pollock, J. (1987). Defeasible reasoning. *Cognitive Science, 11*, 481-518.

Thagard, P. & Nisbett, R.E. (1993). Variability and confirmation. In R.E. Nisbett (Ed.) *Rules for Reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Touretsky, D., Horty, J. & Thomason, R. (1987). A clash of intuitions: The current state of non-monotonic multiple inheritance systems. *Proceedings of IJCAI-87*, 476-482, Milano, Italy.