AI Magazine Volume 11 Number 5 (1990) (© AAAI)

# Workshop on Defeasible Reasoning with Specificity and Multiple Inheritance

*R. P. Loui*

A workshop on defeasible reasoning with specificity was held under the arch in St. Louis during April 1989, with support from AAAI and McDonnell Douglas, and the assistance of Rockwell Science Center Palo Alto and the Department of Computer Science of Washington University.

The workshop brought together proposers of systems of non-monotonic or defeasible reasoning that exhibited subclass or specificity defeat. There were twenty invited participants. The program committee (David Etherington, Hector Geffner, and David Poole) also invited an equal number of participants from those responding to the call for participation. One third of the attendees of the workshop came from abroad.

Twenty-six participants submitted three-page notes on their current thinking, and there were synopses of a dozen existing formal systems. These notes were edited by Michael Kahn and were available at the workshop, but, by agreement with the authors, will not be distributed.

As advertised, the workshop was recorded. An edited version of the recordings can be obtained by contacting the workshop organizer (this author). The procedings have been transcribed and efforts are being made to edit the transcripts and make them available.

The workshop was supposed to provide a venue for challenges to each system and to allow the airing of disputes already on record. However, it was discovered that many of the disputes are no longer pointed. In controversy's stead there was a free exchange of ideas, and the development of more general perspectives on emerging work. Review of the tapes shows a dense, rich exchange, especially on methodology.

The workshop program consisted mostly of panels. There were also provocations by persons with extreme positions, a poster session, and a problem session for working various problems with various systems. Surprisingly, there was little concern over which systems solved which problems. The workshop was essentially a series of partly planned presentations: each person associated with a system or a contribution had twenty minutes to explain himself to a community that needed no background and would countenance no salesmanship. The pace was frantic, with tightly scheduled activities consuming all but eight hours a day. By weekend's end, all participants were exhausted.

Peculiar to this workshop was the participation of some senior philosophers of science and philosophical logicians, who were able to lend perspective. The workshop also contained previews of several of the remarks heard at the First International Conference on Principles of Knowledge Representation and Reasoning (KR89), held in Toronto during May 1989, including remarks by Kautz and Selman, Doyle, Pearl, Etherington, Neufeld, and Poole.

The workshop began with a session titled "What is this thing we're trying to formalize?" David Etherington noted there is temptation to say that Geffner's system (in *Knowledge Representation and Defeasible Reasoning.,* H. Kyburg, et al., eds., Kluwer 1990) just is the kind of defeasible reasoning we've all been trying to formalize. But, he reminded us, that's what we thought of Touretzky's system for defeasible inheritance in 1985. Ron Loui tried to establish a convention regarding notation in inheritance hierarchies and couldn't. (Thereafter no subsequent attempt to standardize syntax or vocabulary was made by anyone.)

Donald Nute reminded us that defeasible reasoning is not always motivated by probabilistic concerns; L. Thorne McCarty gave an example from British civil law. Judea Pearl thought that the example could be made probabilistic with consideration of utilities. Ben Grosof pointed out that one can always find a utility model under which adoption of a rule is justified. Nute explained objective criteria other than truth conditions and justification conditions for adopting rules, including assertability conditions and compliance conditions. One can comprehend the conditions for complying with a rule, e.g., a maxim for chess playing, without knowing the conditions for its justification. David Poole thought that this supported his proposed requirement that inventers of systems write user's manuals.

Pearl felt that the disputes about system behavior had to do with behaviors on which we lack strong intuitions. He wondered if we could conceive of an era when decisions on the esoteric questions could make a drastic difference.

Six impromptu talks were given the first evening in parallel with a problem session that elicited discussion on some fifty benchmark problems that had appeared in the literature. A list of these problems is available upon request.

The next morning's session focused on research methodology. Poole claimed that reasoning is based on arguments; people reason amongst themselves in this way; his program, THEORIST, tries to be the simplest argument system. He claimed that he just uses logic; if you don't like the conclusion, criticise the premises. Kurt Konolige ferretted out the admission that much of this logic occurs at a meta-level. This deliberate conflation of logic and meta-logic agitated the audience at KR89, where the remarks were reiterated.

Fahiem Bacchus professed a purely statistical view of defeasible reasoning. John Pollock, Konolige, and Etherington were concerned that defeasible reasoning underlies the statistical reasoning to which Bacchus appealed. (Grosof had given a poster talk on this point the night before.) Bacchus thought there was room to call rules for selecting reference classes policies, rather than reasoning; this is a long-standing view of Henry Kyburg.

Jim Delgrande explained that in his system, default instantiation relies on an assumption of least exceptionality of the world. Poole felt that if you're willing to make assumptions, you don't need conditional logic to define possible worlds in the first place.

Jeff Horty displayed an inheritance net and asked what conclusions each of us drew from the net. When asked what the links meant, Horty said they

meant something like tendency. In one of the sharpest exchanges, Shastri pressed on the meaning of tendency: "You don't tell me what the links mean then you ask me what conclusions to draw!" Horty quipped, "Sometimes that happens." Horty claimed rhetorically that there were exactly 72 possible theories for inheritance and contemplated the translation of each to classical non-monotonic logics. Matt Ginsberg asked if the path-based work would go away if the translation were successful. Horty thought they would, except for implementation. Rich Thomason thought that translations reminded him of model theory for proof theories.

The most voluble session asked the senior philosophers to comment on the role of convention in the design of logic, and on defeasible reasoning's place in philosophical logic's future. David Israel introduced the speakers, but first remarked that inheritance was supposed to be simple and had nothing to do with problems such as the Yale Shooting Problem. Perhaps we ought not to talk about all of defeasible reasoning at once.

Jim Fetzer claimed that defeasible reasoning is inductive reasoning when you don't know what you're doing. He reviewed Hempel's framework for scientific reasoning and asked us to clarify things in this framework that are unclear in defeasible reasoning. Examples include the purpose of inquiry and the seriousness of making mistakes.

Henry Kyburg thought it's not going to be easy to decide whether or in what sense the world dictates one logic or another. He listed for our consideration classical, temporal, modal, deontic, causal, intuitionistic non-monotonic, default, inductive, and probabilistic logics—all are in the same boat. Then he suggested that there are nevertheless principles for choosing among conventions: not semantic basis nor intuitive persuasiveness, but simplicity, power, familiarity, elegance. Kyburg also felt that specificity construed as subsets is inadequate—specificity taken as logical strength is just the total evidence requirement of induction.

Pollock told us not to worry about semantics until we have a better idea what the target is. He also noted that theoretical reasoning is skeptical whereas practical reasoning is credulous; this eliminates one of the problems in Touretzky et.al.'s "Clash of Intuitions" (IJCAI 1987). These are similar to the remarks he made at the AAAI Spring Symposium.

Nute added to the assault on semantics, stating that Frege and Russell probably weren't worried about completeness "when they were working on what we now know as God's logic [first-order logic]."

Thomason pretended to be a computer scientist and recommended keeping the philosophers at arm's distance. He felt that philosophical logic now belongs in computer science and that most of the former's new directions will come from the latter. Philosophical imperialism is bad for us; conventionalism can't be right—we can't just go around choosing our logic. Fortunately, the needs of users force us to keep our feet on the ground. Thomason joined Poole in requiring that manuals for systems be written. He also suggested that someone try to put defeasible and probabilistic reasoning together in one system.

The afternoon began with a session on defeasible reasoning and probability. Ben Grosof, as moderator, explained the view that inheritance arises in probabilistic reasoning, so probabilities can't be the underlying semantics for defeasible reasoning. This idea had been mentioned earlier by several people and Grosof clarified it.

Bacchus tried to defend a system in which only a single defeasible link can be used in any chain of reasoning.

Eric Neufeld defended his system, motivated by qualitative probabilistic relationships, which reasons about shifts in beliefs. The topology encodes an underlying probability distribution that is factorable like influence diagrams, that licenses inferences that are not in other systems. For instance, it is reversible for diagnosis. Ginsberg found it counterintuitive, stating that it was "an attempt to model my reasoning in a certain mathematical way that has as a result something I find totally counterintuitive."

Neufeld also discussed Simpson's paradox and the lottery paradox in the context of defeasible reasoning. Pearl thought it was a great example of convention versus probability.

Hector Geffner talked about the conservatism of Ernest Adams' rules (in *Aspects of Inductive Logic,* J. Hintikka and P. Suppes, eds., Elsevier 1966), noting that they are shared by Lehmann, Delgrande, Makinson, and Maggidor. He explained that the last rule in his system assumes that the antecedent of a defeasible rule holds in the presence of other evidence. This raised a discussion on irrelevance, and justifying extensions to the core inference rules that are widely shared.

John Pollock was then invited to provoke the audience. Pollock felt that defeasible reasoning has a more complicated logical structure than has been appreciated in inheritance. Specificity defeaters are an incomplete generalization of subset defeaters. "Most penguins do not fly" does not entail "Most birds fly," and both are legitimate parts of the antecedents of reasons, in the way that Pollock writes reasons. Pollock also voiced the need for a projectibility constraint to prevent inheritance from disjunctive classes, for instance.

Saturday ended with a session on issues and principles. Horty continued to bemoan the numerous choices in inheritance about which he has no preference. He also raised contraposition as an issue of contention. Lin Padgham suggested that contrapositives be added as rules at a lower priority after drawing primary default inferences. Everyone seemed to agree, though there was a split among those who wanted it and those who did not.

Ginsberg debunked a naive argument against contraposition. He went on with an attack on anyone who would describe inheritance reasoning in a complicated way in their frameworks just to promulgate their tools.

Poole focused on the distinction between background and contingent evidence, noting that it arises in probabilistic approaches and path approaches, as well as in logic-based approaches like THEORIST.

Loui tried to carve a distinction between two paradigms for defeasible reasoning systems: irrelevance-based, which has axioms and a monotonic proof theory, and argument-based, which is always done in the meta-language. Pearl summed it as "Persons born inhibited need to be encouraged; if you are promiscuous, then your behavior needs to be limited." But Loui felt the two paradigms behave differently under computational limitations. The session ended abruptly during musings on the relation between irrelevance and lack of defeat.

Sunday morning began with a discussion that was supposed to be on path-based versus model-based approaches to defeasible reasoning. Gerhard Brewka told us that defeasible reasoning can be construed as inconsistency-handling. He enjoined us to compute (preferred) maximal consistent subsets, since we have to solve the problem of handling inconsistencies anyway: in this approach, specificity is implicit.

Geffner felt that a focus on inheritance means commonalities are obscured. Many solutions to our problems have been understood only in terms of particular systems — inheritance is the biggest offender. For example, irrelevance is common to defeasible reasoning systems.

Michael Gelfond argued for auto-epistemic logic and reductions thereto, because he believes most formalisms reduce to auto-epistemic logic. He complained that not to translate systems makes it hard to see mathematical properties of new systems. Etherington wondered whether the exercise of translating inheritance into auto-epistemic logic tells us anything about inheritance or about auto-epistemic logic. Ginsberg reiterated his point that there are general lessons to be learned in a general framework, which are lost by translating to special frameworks, e.g., preference of more premises over fewer. He asked if Gelfond believed in that, and Gelfond did not.

Brian Haugh depicted arguments as just richer relations between antecedents and consequents than links. Also, model-based theories with specificity have problems: apparently, successful theories must refer to links or arguments—they must refer to syntactic objects.

Sunday continued with remarks on implementation and complexity. Henry Kautz was aghast that there would be a separate session devoted to these topics. AI doesn't take computation as a metaphor, he inveighed, but takes thought as a specific kind of computation.

Randy Goebel related a story about Theorist's development: people said it couldn't be done but they had it up and running. He also felt that hands-on work would simplify defeasible reasoning systems more than continued mathematical work.

Thomason noted that the main reason to be interested in inheritance is its close connection with computationally effective deliverables.

Lynn Stein argued that tractability is irrelevant because formalism is just now being produced. We still have no agreement on what the right answer is, which is typical in commonsense reasoning. So we say of research like Touretzky's that its contribution was a specification that was more right than shortest path, if still not completely right. Its NP-completeness was not what made it a contribution. Grosof recommended to a weary audience that "perhaps we should inherit this discussion from AI in general."

Padgham then discussed the implementation of her inheritance system. Finally, Bart Selman discussed his various tractability results from collaboration with Levesque and Kautz.

Jon Doyle was asked to provoke the crowd. His work with Mike Wellman indicated that the search for a unified logic to resolve preferences in desirable ways is impeded by Arrow's Theorem, made famous in economics and social choice theory. One possible way out is to limit the domain of preferences. Etherington, Grosof, and Konolige all cried that the pessimistic result would not apply to skeptical reasoners? Doyle felt that even if it did not apply to skeptical reasoners, it applies to choice between skepticism and credulity. McCarty suggested that we can learn from what happened in social choice theory. The ideal is unattainable, but in lieu of the ideal, we can and should ask how to cope with the impediments to achieving the ideal.

In the summary session, Etherington was still concerned that we couldn't distinguish progress from motion, and that we had agreed on no explicit requirements for future papers. Konolige talked about having more "cribs in the deck," i.e., places to file the various defeasible reasoning systems that have been developed, in an orderly fashion. He felt that there weren't any more general principles to be found and that domain-dependent, flexible specification of priorities was the next area to investigate. Konolige also brought up the idea of resource bounds.

Thomason related his experience with the plethora of modal theories, where there are just as many alternatives as in defeasible reasoning; people learn to live with the alternatives. Theoretical progress has been really remarkable in defeasible reasoning, said Thomason, but he again reminded us to connect responsibly with "technology," that is, with implemented systems. A lot of the problems might also disappear if we were to focus on decision-making. Generally, he pled for unity.

Doyle reminded us that basing our systems on intuition is fine if we are willing to live with intuitions that differ between people and change over time. He felt that the underlying intuitions invoked economic issues. Earlier, he had referred to social choice theory; here he mentioned the costs of reasoning under limited resources. Are there 72 theories, as Horty says, or 72 different utilities?

David Israel then led a discussion on limited rationality. He linked it to the pressures to commit to belief, to defaults, and to intentions. Finally Israel took a larger view to assuage some worries. He asked us to imagine what might have happened if complexity theory had arisen before 1928. He thought that even without "the crazy stuff about semantics," paths were very good for inheritance, and arguments were very nice for defeasible reasoning. He recalled that in 1957 there was a symposium titled "Is There One Correct Modal Logic?" which seems ludicrous today. But he warned, there might yet be a difference —it was obvious that all the modal logics invoked different notions. Our clashes over defeasible reasoning somehow seem more substantive.

Overall, the workshop was a success despite its exhausting nature and the fact that it did not achieve its original goals.

More worrying is that the published output in this area has diminished since the workshop. Perhaps this is because the workshop settled many of the disputes out of print. It established an understanding among the active researchers of what would be considered old and new. Future papers should be better for avoiding methodological issues that are better recorded in a workshop like this.

## About the Author

R. J. Loui is Assistant Professor of Computer Science and Adjunct Assistant Professor of Philosophy at Washington University in St. Louis. He currently devotes his time to expanding the idea of competing arguments in formal reasoning, beyond AI, and (he hopes) beyond this decade.