Washington University in St. Louis

Washington University Open Scholarship

All Computer Science and Engineering Research

Computer Science and Engineering

Report Number: WUCS-93-39

1993

Analogy, Decision, and Theory-Formation as Defeasible Reasoning

R. P. Loui

The development of computationally informed formalisms for reasoning with defeasible rules affords new accounts of familiar forms of reasoning. This paper points to recent accounts of defeasible reasoning and portrays analogy, decision, and theory-formation as essentially defeasible, in the same way that statistical reasoning has been portrayed. Each portrayal depends largely on the idea of partial computation, which is inherent in actual reasoning, largely ignored by past formalisers, and formalizable now.

Follow this and additional works at: https://openscholarship.wustl.edu/cse_research

Recommended Citation

Loui, R. P., "Analogy, Decision, and Theory-Formation as Defeasible Reasoning" Report Number: WUCS-93-39 (1993). *All Computer Science and Engineering Research.*
https://openscholarship.wustl.edu/cse_research/534

Department of Computer Science & Engineering - Washington University in St. Louis
Campus Box 1045 - St. Louis, MO - 63130 - ph: (314) 935-6160.

Analogy, Decision, and Theory-Formation as Defeasible Reasoning

R. P. Loui

WUCS-93-39

August 1993

Department of Computer Science
Washington University
Campus Box 1045
One Brookings Drive
St. Louis MO 63130-4899

Submitted to SYNTHESE May 1994.

Submitted to *SYNTHESE 5/94*

ANALOGY, DECISION, AND THEORY-FORMATION AS DEFEASIBLE REASONING

R. P. Loui
Department of Computer Science
Washington University
St. Louis, MO 63130

ABSTRACT.

The development of computationally informed formalisms for reasoning with defeasible rules affords new accounts of familiar forms of reasoning. This paper points to recent accounts of defeasible reasoning and portrays analogy, decision, and theory-formation as essentially defeasible, in the same way that statistical reasoning has been portrayed. Each portrayal depends largely on the idea of partial computation, which is inherent in actual reasoning, largely ignored by past formalizers, and formalizable now.

1. INTRODUCTION.

1a. DEFEASIBLE REASONING

Defeasible reasoning refers to non-demonstrative reasoning where the reasoning makes explicit use of rules that admit of defeasance (other authors have used the words exception, preemption, preclusion, interference, rebutment, and defeat). The use of the term "defeasible" originates with the great legal philosopher and Austin-Ryle student, H.L.A. Hart, who apologizes for importing it from English property law. Defeasible reasoning dates to antiquity but formalism has appeared only in recent decades, especially beginning with Pollock, who follows Chisholm, and Kyburg, who responds to Reichenbach. Nute, Glymour-Thomason, and Belzer are other philosophers who have produced formal systems. Most of the work has been done within artificial intelligence (AI) by a large number of authors, especially beginning with Reiter, McCarthy, Moore, McDermott-Doyle, followed by Sandewall, Touretzky-Horty-Thomason, Loui, Poole, Geffner-Pearl, Vreeswijk, and others. Defeasible reasoning is the formal characterization of what has long and wide been acknowledged as ceteris paribus reasoning.

Frequently no distinction is made between defeasible reasoning and non-monotonic reasoning in AI, but a distinction can be made. The former requires defeat among rules, that is defeasible rules. the latter refers to how theorems grow as axiom sets are varied: L is *non-monotonic in syntax* just in case for some sentences S and T, $Thm_L(S)$ is not contained in $Thm_L(S \cup T)$. Non-monotonicity need not arise from defeat of non-demonstrative rules. For example, there are paraconsistent logics, the recent step-logic of Elgot-Drapkin, Perlis, and Miller, and the minimization of models following John McCarthy. In these non-monotonic systems, non-monotonicity cannot be attributed directly to rules being defeated. Nevertheless, non-monotonic reasoning is most often viewed as arising from rules that are defeated under certain conditions.

The most important (and controversial) part of defeasible reasoning is the specification of conditions under which a rule is defeated. A rule, if P then (defeasibly) Q, is defeated by a more specific rule, if P and R, then (defeasibly) not-Q. This is the so-called "implict specificity" defeater that some have rejected, but which is irresistable to others. It is also defeated in the presence of indefeasible knowledge that not-Q, about which there is no controversy. Agreement over rules for defeat has been elusive. Researchers have accepted the inevitable "clash of intuitions" (the name given by Touretzky, Thomason, and Horty) and adopted a conventionalist stance on language design. Most authors seem willing either to (a) adopt defeat conditions apropos to a particular form of reasoning, e.g., legal reasoning, taxonomical reasoning, reasoning about action; or (b) to ascend to meta-argument about which argument is better; or else to specify priority among rules explicitly. In legal circles, for example, lex superior, lex posterior, explicit subordination, and meta-argument are all recognized in addition to specificity, which they call lex specialis.

Uses of defeasibility that are already familiar to philosophers are in practical reasoning and ethics and epistemology. There, defeasible reasons are ceteris paribus or prima facie reasons. In ethics, defeasibly do action A because it achieves P which is good, but do not do A because it achieves P and Q, which taken together are bad. In epistemology, the appearance that P is defeasible reason for P.

Artificial intelligence has sought defeasibility for two purposes: convenience and delayed commitment:

(1) Defeasibility is unavoidably convenient for specifying information in databases. This may include a database's information about how to project the effects of actions into the future in an automatic reasonaer about plans. It may include assumptions about the linguistic conventions about names (e.g., unique names name unique objects), object properties (e.g., properties, if not asserted to be true, are false), and existence of events (e.g., no events occurred unless mentioned).

(2) It permits policies to be specified in databases without requiring advance knowledge, commitment, or computation of what the conclusions or actions the policies imply. This includes legal and other policy reasoning, reasoning about past cases to construct new solutions to problems, non-deterministic resolution of conflicting rules in expert systems. It also includes, this author claims, many of what are considered purposes of the first kind, such as reasoning with the rule "birds fly" or the rule, "marry young."

One reason why AI has been forced into defeasible and non-monotonic systems is that limits on computation are real concerns. Real problems cannot usefully be abstracted under an assumption of ideal computation. One concern is finiteness of computation. When the underlying logic is semi-decidable, attempted proofs of non-theorems can be non-terminating. Semi-decidability has led to "negation as failure," the practice of assuming the negation of P if P cannot be proved. Typically, the concern is simply to reach conclusions quickly, whether the conclusions are decidable or not. Limiting the amount of time allowed to check for defeat, exception, or counterargument converts defeasible rules used for the first kind of purpose into rules of the second kind.

Current styles of formalizing defeasibility either introduce a new connective in the language, for defeasible conditionals, or else start with the metalinguistic relation of one sentence being reason for another sentence.

In the former approach, the sentence 'P --> Q' is well-formed in addition to 'P ==> Q'. The result is like a logic of counterfactuals, though the emphasis is on proof theory instead of model theory, and almost no axioms are shared with conditional logics in mainstream philosophers' ken. Delgrande and Lehmann try their best, however, to stay close to such as Stalnaker and Lewis. Cross and Boutilier try to relate their work to Alchourron-Gaerdenfors-Makinson revision and Ramsey's rule. Many authors envision game-theoretic semantics (e.g., Hintikka), or else follow Nute and Pollock and defend their right to provide meaning operationally without equivalent mathematical modeling.

In the latter, the sentences 'P' and 'Q' are related, 'P' >-- 'Q'. A defeasible entailment relation, |~, is defined using the existing deductive entailment relation, |--. This is done by defining objects such as arguments or defeasible derivations, then defining relations between the objects, such as disagreement between arguments or defeat of one argument by another (Kyburg, Loui, Simari-Loui, Geffner-Pearl, Sartor, Prakken, and Vreeswijk). In these approaches, defeasible conclusions are held at a lower pedigree or stratum than that from which they were concluded. This semantic ascent, ascent to the metalanguage, does not introduce new model-theoretic dilemmas, but it does revive the problem of reconciling assertion in the object language with defeasible warrant in the metalanguage.

An older style attaches explicit censors to each rule, which, if taken to be true, would defeat the rule. This forces a definition of defeasible entailment that refers to fixed points, or equilibria (Reiter, McDermott-Doyle, Moore, Touretzky). These enumerated styles still do not exhaust all approaches: the system based on multi-valued logics (Ginsberg) is not properly described as any of the above.

Defeasibility of rules can be regarded in two ways. The way in which defeasibility is regarded is more important here than the formal mechanism in which defeasible rules are embedded.

Some regard defeasibility as a style of specifying doxastic commitment that could be specified in some other way, albeit less conveniently. Defeasibility as a style has advantages because it permits generalization from near-regularities in addition to perfect regularities. "Birds fly, but Tweety doesn't." Some regard this as a succinct report on the flying of all birds, Tweety and all others. There may in addition be information here about how to revise the commitment in a few situations. For instance, during expansion to include the belief that Opus doesn't fly, there is a record of the grounds upon which defeasibly one concluded Opus does fly. Defeasiible specification of knowledge frequently makes a commitment to a belief revision strategy in addition to commitment about current belief.

Defeasibility has also been viewed (Pollock, Doyle, Loui, but consider Etherington-Reiter, Toulmin, and earlier traditions) as a way of specifying how beliefs can be constructed under partial computation. That is, defeasible rules are potential inputs to a non-demonstrative process of constructing belief. Implicit in this view is that construction is non-ideal, if there is an ideal computation. However, certain kinds of constructions may still be valuable. The computational concern here is not directly with computational complexity classes such as polynomial-time. Rather, the concern is more mundane: computational bound forces warrant to be defined with respect to subsets of the data available, or more precisely, subsets of the arguments constructible. More computation is better, regardless of asymptotic behavior, and regardless of pathological cases.

These two views are briefly explored in the next two subsections.

1b. DEFEASIBILITY AND LINGUISTIC STYLE

Adopting defeasible rules alters the logico-linguistic environment. It is tempting to say that the rules, e.g., 'P is reason for Q,' 'P & R is reason for not-Q,' cannot be replaced with any combination of material conditionals or counterfactual conditionals. This irreplaceability spurred interest in non-monotonicity among AI researchers at first. This is not quite right if one adopts non-computational purposes for defeasibility; it is correct if the purpose is constructive.

Consider the commitment represented by the defeasible rules together with the indefeasible assertions (also called the evidential basis, the assumptions, the manifest beliefs, etc.). The commitment to 'P is reason for Q,' and to P, by itself, is also a commitment to Q. Meanwhile, a commitment to 'P is reason for Q,' and to 'P & not-Q,' does not also contain a commitment to Q. Thus, a set of defeasible rules together with the assertions upon which defeasible conclusions are drawn defines a set of theorems in the (defeasible) logic of the language adopted. This set of theorems is fully determined by two things: (1) the rules and (2) the assertions. So the doxastic commitment can equally be specified by listing just the theorems. Tthis is like an elimination lemma, but the observation here is not profound because finiteness of the resulting specification is unlikely. A function that maps assertions to theorems can be created with defeasible rules that cannot be recreated with any set of demonstrative or counterfactual conditionals. But this is not too important: theorems are still fully determined by rules and assertions; hence, once the commitment is fully written, it can be rewritten in an indefeasible way.

The defeasible rules also represent a commitment to a belief revision strategy. This ''confirmational'' commitment has been lost in the elimination just envisioned, but it too can be specified without resort to defeasible rule. For example, index the sets of indefeasible assertions by the sentences whose adoption would force a revision, as do Kyburg, Levi, Gaerdenfors, and others. Let $<R, K>$, where R are rules and K are assertions, be replaced by K_0 = DefeasibleEntailments(R, K), and let $K_{<A>}$ = DefeasibleEntailments(R, K u {A}), for all A. The set of $K_{<A>}$'s is the equivalent commitment.

A knowledge representer chooses language, axiom, and assertion all of a piece. AI discovered that the selection of an automated reasoner, a choice of language, is as easy as (in fact, easier than) the selection of its input, a choice of assertions. It thus becomes appropriate to view commitment as arising from all three in concert. It becomes appropriate to compare two representations of commitment by considering the total. Adopting the assertion P alters the logico-linguistic environment, changing what the next assertion would conspire with prior assertion to represent. Adopting the rule 'P is reason for Q' similarly alters the logico-linguistic environment. Thus, commitment can be specified both in the object language and in the metalanguage, whereas commitment is usually specified just in the object language.

To portray familiar forms of reasoning as defeasible reasoning, and to claim that something new is thereby achieved, a different view of defeasibility is required.

1c. DEFEASIBILITY AND COMPUTATION.

There is the other view of defeasible reasoning in which rules are not merely a commitment that can be rewritten, but are instead rules that govern the process of constructing belief. Defeasible reasons constrain the arguments that can be produced in deliberation. Conversely, they also license the production of well-formed arguments. Deliberation must follow a protocol according to which argument and counterargument are produced with ample opportunity: though resource-bounded, the process must nevertheless be fair and efficient. The rules participate in a fair, efficient process to produce rational belief.

Rational belief is a relation, not a function of what is written. Several different sets may be related to manifest commitment as the defeasible conclusions thereof, and those sets might not contain one another. In a resource-bounded classical theorem-prover, the conclusions grow monotonically with computation. Here, instead, the growth is *non-monotonic in computation.* The process of drawing defeasible entailments from a fixed set of rules and assertions might cause contraction as well as expansion of beliefs. Describe rational belief as a function of several things: rules, assertions, and the parameters of a process that constructs belief. Parameters might specify resource constraints, information about the pragmatics of deliberation, and even non-deterministic choices. In the latter case, the process that is constructing belief is regarded as a non-deterministic computation. The language-user who writes commitment to defeasible rules understands that part of the rules' meaning is that they can be used in this way. Calling such processes "non-ideal" is a misnomer since the appropriateness of a belief may be defined in terms of the process that constructed it.

This view may be familiar to constructivists, intuitionists, and dialecticians and is anathema to the Russellian, Fregean, Carnapian logical investigation of reasoning and science. The view has numerous historical precedents, none of which have been mainstream in philosophical logic or philosophy of science in this century.

It is correct to think of this as procedural rationality: features of the procedure make the result rational, not just the results' substantive relation to the procedure's inputs. Adherents to the procedural rationality concept usually think of heuristic search, satisficing decision-making, or complex programs whose organization is impenetrable. In contrast, the process envisioned here is simply the process of constructing arguments from conditionals of the form "if P then defeasibly Q", and the construction of counterarguments, rebuttals, and so forth. This idea resonates in the writings of Herbert Simon, Nicholas Rescher, and earlier writers such as Popper and Keynes who took dialectic and juridical reasoning seriously. Thus, computationally bounded processes construct ideal procedurally rational beliefs though they do not construct ideal substantively rational beliefs.

It is useful to think first of policies (legal rules, social conventions, aphorisms), which clearly are defeasible rules of this kind; then, to extend the metaphor to include rules of reasoning and specification of doxastic commitment. It is not useful to insist that policies be reduced to something probabilistic or decision-theoretic. Certainly policies are adopted for various reasons: reasoning with a policy might properly involve other forms of reasoning, such as decision-theoretic reasoning or inductive reasoning. Some policies are surely adopted on statistical grounds. However, irrespective of whether policies can always be reduced to the reasoning that led to their adoption (which is a question that belongs to legal scholars for legal rules, to social scientists in the case of social conventions, and so forth), there is a non-demonstrative method of processing defeasible rules as policies. It is this kind of process that also can construct belief from defeasible rules.

Traditional studies view rational belief as imposition of constraints, where rules provide constraint. Traditionally, language use is restricted so that it cannot violate the constraints. The alternate view, taken here, is that rational belief as the result of non-deterministic process guided by rules.

Two properties distinguish constructive processes from constraint-propagating restrictions on the use of language. The distinguishing properties are (1) non-determinism of outcome and (2) non-monotonicity in computation.

The process that constructs a rational belief might not have constructed the belief that it did; worse, it might have constructed the contradictory belief. Had the process been permitted more computation, more time, a different conclusion could have emerged. Warrant is conferred anyway because of the way in which it was constructed: the process was fair, efficient, and appropriate.

Legal hearings, political deliberations, and binding arbitrations are examples of these kinds of processes. So are sports championship playoffs, Senate confirmations, and academic examinations. So are the computations of programs that reason. So, too it seems, can be inquiry that uses statistical, analogical, decision-theoretic, and best-explanation reasoning.

2. FOUR KINDS OF DEFEASIBLE REASONING.

Four kinds of reasoning are presented as varieties of defeasible reasoning. In the first three, defeasibility could arise because information is added through additional computation (internally), or because it is added through experience (externally). In the last, the defeasibility must be the result of further computation or deliberation upon fixed evidence. This section of the paper seeks to show how defeasibility of the computational kind enters important forms of reasoning, and how analysis of these kinds of reasoning can apply defeasible reasoning. In each case, a much deeper analysis could be given, and other work seeks to do this. Here, it suffices to make the case that an analysis based on defeasible reasoning captures an important aspect of the kind of reasoning.

To traditional students of non-deductive inference, the following will at first seem too simple. The schemata presented are the prima facie reasons, the defeasible reasons. The complexity of the pattern of inference arises from the competition among the many defeasible reasons. In statistical reasoning, for example, it will at first seem that the schemata pay no respect to the total evidence requirement. A statistical inference should use the totality of available evidence. The defeasible reason permits a line of argument based on subsets of the evidence of any small size. In analogical reasoning, it will seem that there is no place for distinctions among source and target. There seems to be no place for multiple conflicting cases. The schema provides locally only for argument and not for counterargument. It is the aggregation of the arguments pro and con that produces the expected complexity. Specificity drives computation toward arguments that use the totality of evidence. The rules regarding reinstatement of arguments against counterargument force analogical arguments to take account of distinctions between source and target.

2a. DEFEASIBLE STATISTICAL REASONING.

The case for regarding certain forms of statistical reasoning as defeasible reasoning has been ably made by John Pollock over the past decade.

The reasoning in question is the inference that an event has a probability by appeal to a reference class. For example,

STAT 1 (basic schema)
the frequency of H's among A's is k_1;
t is an A;

---

thus, defeasibly,
the probability that t is an H is k_1.

Given that the frequency of sports car drivers having an accident in a year is 80%, the prima facie probability of a particular sports car driver having an accident in a year is also .8. This inference is defeated by reference to a more specific class. If the frequency of accidents in a year among sports car drivers with good recorrds is 30%, then the probability must be based on the narrower reference class.

STAT 2 (defeater)
the frequency of H's among (A & B)'s is k_2;
t is an (A & B);

---

thus, defeasibly,
the probability that t is an H is k_2.

This is defeat in the strong sense, where one argument deprives another of its ability to interfere with other arguments as well as its ability to support its own conclusion. An example of mere interference would be the following argument, which is not more specific than the first argument (STAT1), nor vice versa. So neither conclusion survives, if STAT1 and STAT3 are the only constructed arguments.

STAT3 (counter)
the frequency of H's among B's is k_3
t is a B;

---

thus, defeasibly,
the probability that t is an H is k_3.

Pollock's account of defeat and interference is slightly different. Instead of supposing that there are two interesting relations (defeat and interference) that may hold between two arguments X and Y, he supposes that argument X and argument Y defeat each other, but there may be a third argument Z which states that X has the appropriate relation to Y (e.g., using a more specific reference class). Z defeats Y, thus breaking the symmetry.

There may be other conceptions of defeat appropriate for this kind of statistical reasoning, and the main intellectual task is to specify these conceptions precisely. But the main point here is that the form of reasoning is captured by introducing defeasible reasons.

What belief is warranted depends on what arguments, counterarguments and rebuttals there might be. There can also be reinstatement, where an argument defeats a would-be counterargument of some initial argument, thus reinstating the initial argument. Consider the argument

STAT4 (extended argument)
the frequency of H's among A's is k_1;
t is an A;

---

thus, defeasibly,
the probability that t is an H is k_1.
$k_1 > l$, the threshold for acceptance based on high probability;

---

thus, defeasibly,
t is an H;
the frequency of G's among H's is k_4;

---

thus, defeasibly,
the probability that t is a G is k_4.

This argument requires a threshold probabilistic acceptance rule (see Swain), and it is legitimate to wonder whether acceptance is defeasible (it appears to be, but see Levi). In either case, suppose STAT4 counters an argument, STAT5, that the probability that t is a G is, say, k_5:

STAT5 (counter to extended argument)
the frequency of G's among C's is k_5;
t is a C;

---

thus, defeasibly,
the probability that t is an G is k_5.

STAT5 would be reinstated by an argument that defeats STAT4. STAT2, for instance, is such a reinstate. STAT2 defeats a subargument of STAT4; it argues that the probability of t being an H is k_2, not k_1, by reference to a more specific class. Of course, the putative reinstater, STAT2 in this case, must itself be undefeated.

An argument that is undefeated with respect to one set of arguments might be defeated in an enlarged set of arguments. We do not suppose a priori that all arguments need be constructed in order for there to be warrant.

Kyburg, who gives a system almost exactly of this form, requires the relevant set of arguments to be the set of all constructible arguments. Probability is an objective matter and a determinate thing. Proper reasoning about probability requires that all arguments be taken into account. Preferring more specific reference classes is one way of imposing the total evidence requirement. It is a matter of style, not a way to make sense of situations in which less than the totality of evidence is brought to bear on probabilistic judgement.

Here, the alternative is to relativize the reasoning to the arguments that have so far been constructed: that is, to relativize warrant to the computation, to the history of the process. A conclusion is warranted just in case an argument for the conclusion is undefeated among arguments that have been constructed. This is not the odious subjectivity that permits any disagreement among individuals. Two individuals who agree about what arguments have been advanced must agree on what conclusion is warranted. However, what is concluded at what times depends on the order in which arguments are constructed, and the resources, the time, available for constructing arguments: in short, on the protocol for argument.

I find Pollock uncommitted on this point; he writes both about ideal warrant and defeasible, non-ideal warrant, which approximates the ideal. Pollock seeks to take defeasible reasoning more seriously than Kyburg, describing processes in which reasons are produced incrementally. The next step toward a more seerrious constructivism is to see constructed belief as ideal by the nature of the process that constructed it, without reference to unbounded cases in the ideal or at the limit.

The process can be internal to the agent, not just the process of accumulating evidence as an epistemic agent goes about interacting with the world. Corrigibility of inductive inference has always been acknowledged because inference is relativized to data and data accumulate. Instead, here the presentation of arguments is sequential, non-simultaneous, because making arguments requires computation.

2b. ANALOGICAL REASONING.

Analogical reasoning has been studied informally, in philosophy, cognitive science, and in artificial intelligence. But systematization has been rare (e.g., Niiniluoto). Recent formal investigations with comprehensive historical consideration are the thesis of S. Russell and his work with T. Davies. The ideas here are a simplification of the analogical models for reasoning from precedent in AI and Law (e.g., Rissland, Skalak, and Ashley).

The pattern of analogical reasoning of interest here is

ANALOGY1 (basic scheme, preliminary)
t is an A;
s is an A;
s is an H;
H is 1-projectible from A (i.e., A is suitable basis for analogies about H);

---

thus, defeasibly, (by analogy of t to s)
t is an H.

Most of the logical interest in this form of reasoning has concerned the nature of the last premise. This is the part of the argument that asserts that a property is suitable for projecting another property from a single instance. For example, from a single case we may be willing to project whether a car is expensive or not based on its marque. If a single expensive Cunningham were produced, we will reason by analogy of a second Cunningham that it is also expensive. But we may withhold 1-projectibility of intelligence from birthplace. The first person from Lanikai, who is a genius, does not permit by analogy the next Lanikaian to be presaged a genius. Projectibility is a concern in probabilistic reasoning, too; we suppressed its discussion in the previous section.

Apparently, 1-projectibility sometimes amounts to knowledge, perhaps from some theory, that the frequency of a property in a class is extreme (together with knowledge that one property is projectible from the other in the first place). But which extreme, extremely high or extremely low, is unknown. The exemplar which is the analogical source resolves the issue.

What is interesting here is the variety of ways in which an analogical argument can be attacked, and can be defeated, ignoring the problem of projectibility.

The conclusion that t is an H should defer to background knowledge that t is not an H, if present.

Also, an analogical argument from different properties, to t is not an H, should interfere:

ANALOGY2 (alleged counter)
t is a B;
r is a B;
r is a non-H;

---

thus, defeasibly, (by analogy of t to r)
t is a non-H.

Similarly, if a non-H member of A were discovered, the representativeness of the putative source, s, would be impugned. In fact, the discovery of a non-H member of A affects the original analogy in two ways: first, it questions the connection between H and A suggested by the original source, s; second, it permits a competing analogy concluding that t is not an H.

The improved analysis for analogy, then, is:

ANALOGY3 (basic schema, revised)
s is an A;
s is an H;
A is suitable for analogies about H;

---

thus, defeasibly,
being an A is defeasible reason for being an H;
also
t is an A;

---

thus, defeasibly,
t is an H;

Discovering a non-instance, r, which is a non-H and an A, permits two counterarguments. The first attacks ANALOGY3's subargument:

ANALOGY4 (meta-counter)
being an A is defeasible reason for being an H;
r is an A;
r is a non-H;
A is suitable for analogies about H;

---

thus, defeasibly,
being an A is defeasible reason for being a non-H;
and
it's not the case that
being an A is defeasible reason for being an H;

(where presumably it is the second conclusion that actually qualifies this argument as a counterargument).

The second, attacking the top of ANALOGY3 with a competing, interfering analogy:

ANALOGY5 (counter)
r is an A;
r is a non-H;
A is suitable for analogies about H;

---

thus, defeasibly,
being an A is defeasible reason for being a non-H;
also
t is an A;

---

thus, defeasibly,
t is a non-H.

A more specific analogy would not only interfere with, but also defeat ANALOGY3:

ANALOGY6

part i. (defeater)
r is an A & B;
r is a non-H;
A & B is suitable for analogies about H;

---

thus, defeasibly,
being an A & B is defeasible reason for being a non-H;
and
it is not the case that
being an A & B is defeasible reason for H;

part ii. (extended defeater)
also
t is an A & B;

---

thus, defeasibly,
t is a non-H.

Either a defeating counterargument to ANALOGY3's subargument is used (ANALOGY6, part i), or a defeating, more specific analogical argument to the contrary conclusion (ANALOGY6, parts i and ii), or both.

Perhaps the manners of defeating analogical arguments might be reduced to general principles for defeat based on specificity of rules. This would depend on carefully stating the formal rules that permit, for example, part i of ANALOGY6. It would also depend on how generally stated specificity defeat is stated in the supposed argument system. The candidates for such a general principle are all too complex to express here (see for example, Prakken). For example, in the system of Simari and Loui, using ANALOGY7 instead of ANALOGY6 permits the desired reduction, because ANALOGY8 is more specific than ANALOGY7 according to the Simari-Loui specificity principle. The relevant question here is whether authors ought to seek such reductions and systems general enough to do the reducing, or whether we should content ourselves with enumerating the ways of attacking arguments of each kind of argument.

ANALOGY7 (basic schema, alternate)
s is an A;
s is an H;
A is suitable for analogies about H;

---

thus, defeasibly,
if x is an A then x is an H (material conditional);
also
t is an A;

---

thus,
t is an H;

ANALOGY8 (defeater, alternate)
r is an A & B;
r is a non-H;
A & B is suitable for analogies about H;

---

thus, defeasibly,
if x is an A & B, then x s a non-H (material conditional);
also
t is an A;

---

thus,
t is a non-H;

Finally, the analogical argument that t is an H, (ANALOGY1, ANALOGY3, or ANALOGY7), can be attacked by producing an argument of a different kind. For example, there may be a statistical argument that it is not the case that t is an H (which we have been treating as equivalent to t is a non-H, since defeasible reasoning is usually built on top of a standard 2-valued logic; but this need not have been the case). Or there may simply be an adopted policy that C's tend to be non-H's. An argument, then, that t is a C, could be extended to provide a counterargument to ANALOGY3. And there would be no basis for the direct comparison of the arguments, since analogical arguments and policy arguments are in general non-comparable. In particular domains, however, such as law, where the rationales for policies include cases, this need not be the end of the matter; there may yet be a basis for comparison.

If the counterargument is statistical, then it matters whether the frequency statement on which the statistical inference is based contains the source case in its count. If so, then the analogical argument may be equivalent to a weak statistical argument, and one may in fact defeat the other. However, an equivalence may not be possible if the projection takes into account some special status of the analogical source as a precedent.

Amazingly, inductivists have conceived of "the" probability of an event. In contrast, those who have identified analogical patterns of reasoning write indeterminately about "an" analogy, not "the" analogy. Somehow, induction has been held to an ideal to which analogizing has not also been held. Perhaps for this reason, analogy has been held to be the greater scandal. However, we now find analogy to be the more honest about the non-determinism, process, and computation underlying inference.

2c. THEORY-FORMATION AND BEST EXPLANATIONS

In some situations, inferences are warranted against the background of a theory, which is the best fit to some data, or else implied by our best explanation of the data. Historically, the main disagreements regarding this kind of inference concern what is the criterion of fit, what counts as an explanation, and how explanations are to be ranked. Simplicity and coherence have been proposed as criteria for explanations (e.g., Thagard). Trading predictive power against error of fit is a main consideration in theory-adoption (e.g., Kyburg, whose probability mediates the tradeoff between predictive power and fit).

Suppose that there are fixed and undisputed criteria of fit. There is yet an issue that has been ignored which dominates actual practice of this reasoning, and which can be addressed simply with defeasible reasoning. The issue is the space of explanations and theories, which is liable to be large, perhaps not even recursively enumerable.

Consider (1) the fit of a polynomial to a set of points or the explanatory coherence of a diagnosis with a set of observations of symptom, and (2) the fit of a theory of measurement to observations of relative length or relative preference. Only the first is an example of a space of theories that is plausibly parameterized.

Unless the space of theories can be restricted to a parameterizable set, only a subspace can be searched for best fit. Such restrictions can themselves be regarded as implications of a background theory. Actual claims of best fit must be relativized to the subspace that was actually searched for best fit. A claim of best fit is thus provisional: best among those considered. Inferences that rely on adopting the theory or explanation are thus defeasible, defeated by opposite conclusions derived from the better theory that had not been considered.

BESTFIT1 (basic schema)
T_1 fits data D to degree k_1 by criterion C;
. . .
T_n fits data D to degree k_n by criterion C;
C is the appropriate criterion;
k_1 is greatest among $\{k_1,...,k_n\}$;

---

thus defeasibly,
T_1 is the best fit to data D;
if the best theory is acceptable, and T_1 implies P,
then P.

An argument of this form is attacked by any argument for a contrary of P. Moreover, it is defeated by an argument of the same kind that considers a superset of candidate theories. So it is defeated by:

BESTFIT2a (defeater because more candidates)
T_1 fits data D to degree k_1 by criterion C;
. . .
T_{n+1} fits data D to degree k_{n+1} by criterion C;
C is the appropriate criterion;
k_1 is greatest among $\{k_1,...,k_{n+1}\}$;

---

thus defeasibly,
T_1 is the best fit to data D;
if the best theory is acceptable, and T_1 implies P,
then P.

BESTFIT2b (defeater because more data)
T_1 fits data $D \cup D'$ to degree k_1 by criterion C;
. . .
T_n fits data $D \cup D'$ to degree k_n by criterion C;
C is the appropriate criterion;
k_1 is greatest among $\{k_1,...,k_n\}$;

---

thus defeasibly,
T_1 is the best fit to data $D \cup D'$;
if the best theory is acceptable, and T_1 implies P,
then P.

BESTFIT2c (defeater because conjoined criteria)
T_1 fits data D to degrees $<k_1^1, k_1^2>$ by criteria C_1 and C_2, respectively;
. . .
T_n fits data D to degrees $<k_n^1, k_n^2>$ by criteria C_1 and C_2, respectively;
C_1 and C_2 are appropriate criteria;
$<k_1^1, k_1^2>$ is greatest among $\{<k_1^1, k_1^2>,...,<k_n^1, k_n^2>\}$;

---

thus defeasibly,
T_1 is the best fit to data D;
if the best theory is acceptable, and T_1 implies P,
then P.

There is an issue regarding the computation of the degree of fit. In the case of fitting polynomials to points, the candidate criteria of fit are well known (least squares, number of errors), and computing each is trivial. But in some criteria of fit, determining the degree of fit is a computational problem that may demand a lenient view toward results of partial computation.

For example, Kyburg's plan for adopting universal generalizations in science requires that maximal consistent subsets of sets of first-order logical sentences be compared by size (this computation would be NP-complete even if consistency-checking is constant-time). Thagard's measure of explanatory coherence appears to be defined in terms of an intensive distributed computation with no clear termination criterion. Simplicity and probability may also be involved in the criterion; syntactic simplicity (recall Goodman, Kemeny, and Bunge's discussions; see Kemeny) might be non-trivial, and we have already discussed defeasibility in probability judgement. Determining degree of fit under a complex criterion may be difficult to do in the ideal. Claims about fit will then be provisional on how far the computations were carried. For example, fitting a theory to data could require the rejection of at least k observations, pending consideration of other candidate maximal consistent subsets. Defeasible schemas permit systematization of this kind of reasoning based on partial computation as in BESTFIT4.

BESTFIT4 (schema with partial computation of a criterion based on subsets of D)
Considering the subsets $\{S_{1,1}, ..., S_{1,m1}\}$ of T_1,
defeasibly, T_1 fits data D to degree k_1 by criterion C;

. . .

Considering the subsets $\{S_{n,1}, ..., S_{n,mn}\}$ of T_1,
defeasibly, T_n fits data D to degree k_n by criterion C;
C is the appropriate criterion;
k_1 is greatest among $\{k_1,...,k_n\}$;

---

thus defeasibly,
T_1 is the best fit to data D;
if the best theory is acceptable, and T_1 implies P,
then P.

An interfering argument is superior in at least one of $m_1, ..., m_n$, or n. A defeating argument considers a $<m_1, ..., m_n, n>$ that is not less in any dimension and greater in at least one, reaching a contrary conclusion.

Finally, there is the possibility of counterarguing the premise that C is the appropriate criterion of fit. Until one knows how such a premise can be the conclusion of an argument, one cannot know how such an argument could be defeated.

Thagard attends closely to a process of reasoning to the best explanation. He notes that there is normally theory revision when a better theory is available (whether it is best or not), then proceeds to define a clever enumerative procedure with which all theories can be considered before pronouncing judgement. Thagard's sensitivity to process is oddly paired with his unwillingness to define warrant relative to partial results of the process.

N. R. Hanson's psychological concerns in scientific theory formation could equally well have been computational concerns. Classical theory-formation envisioned the best theory among all possible theories, where all possible theories could be conceived in a single instant. The traditional view was as much a leap as requiring, in deductive inference, commitment to truths that could not be decided in deduction. This leap was consistent, too, with the thinking that produced the total evidence requirement for inductive inference. The objective was to specify ideal commitment. But Hanson thought that there was a logic of discovery, a way of moving from one theory to the next in light of new data. No logic of theory revision, nor of theory discovery is contemplated here. However, the emphasis on ideals which eliminates the role of internal process is attacked here as in Hanson. Traditions in philosophy of science can evolve, now that there are mathematical models for process, and now that computer systems need the guidance of formal theories of rationality, that respect the limits of computation.

2d. DECISION-THEORETIC REASONING.

Since constructive defeasible reasoning provides one form of Simon's procedural rationality, it is no surprise that an account of decision as defeasible reasoning under risk can be given.

The first step, defeasible qualitative reasoning about decision, is easy. Qualitative defeasible decision is just practical reasoning, is natural and familiar. Practical reasoning has long been cast as an exercise in defeasible reasoning. Do A because A achieves P, and P is desirable. This is a reason to perform A: a single defeasible reason is a degenerately short defeasible argument. It is defeated by the argument: don't do A, because A achieves P & Q, and P & Q is undesirable. Similarly, do A, because it achieves P: P is a derived goal, because if P, then doing B achieves Q, which is a primary goal. An interfering argument would be that doing A achieves R, which it is a primary goal to avoid, irrespective of A's enabling B, achieving the desired Q.

Quantitative arguments for decisions can also be cast as part of a deliberative process in which there are argument and counterargument.

Taking an expected utility is a canonical form of arguing for the preferability of a course of action. But it is not the only argument. For example, a multi-attribute model provides an independent way of assigning a measure of preference to a description of an outcome. Utility assessments may be obtained by analogical arguments to benchmark outcomes with incontrovertible utilities. (This is one interpretation of Shafer's discussions of standard lotteries.) Thus, utility assessments of outcome state may be defeasible. In an expected utility argument, the participating probabilities may also be defeasible conclusions.

It is desirable to consider defeasible arguments that argue that a particular state has a particular value, rather than alter the mathematical theory of measurement that underlies the definition of utility. However, instead of lottery comparisons that permit real-valued utility assignments, suppose there is an argument that a particular utility assignment is the right assignment.

Let there be acts, $a_1, ..., a_n$ and events $e_1, ..., e_m$. Value is attached to sequences of acts and events, e.g., $<a_1; a_4; e_5; e_3>$. In an analogical argument about utility, a property, d, is attributed to the outcome denoted by a sequence of acts and events. For instance, d might pertain to something being achieved, or to a monetary consequences of the acts and events. There may be an outcome already given value, in which the property d also holds. On the basis of d-similarity, the two outcomes can be asserted, defeasibly, to have the same value, u. In a multi-attribute argument about an outcome's valuation, the sequence of acts and events is associated with a vector of properties or attribute values. Value is assigned, again defeasibly, because of a regularity among utility valuations that allows a multi-linear functional to be fit.

The pattern of reasoning, for action a_1 and events e_1, e_2 is:

VALUE1 (analogical valuation)

In respect d_1, $<a_1; e_1>$ is like l_1, which has utility u_1;

so defeasibly

$<a_1; e_1>$ has utility u_1,

VALUE2 (expectation valuation)

$<a_1; e_1>$ has utility u_1;

$<a_1; e_2>$ has utility u_2;

prob(e_1) is k_1;

prob(e_2) is k_2;

so defeasibly,

a_1 has utility $k_1(u_1)+k_2(u_2)$;

ANALYSIS1 (basic schema)

a_1's utility is maximum among utilities of acts in $A = \{a_1, a_2, ..., a_n\}$

so defeasibly,

do a_1.

Here, we have supposed analogical arguments supporting the valuations of outcomes. For multiattribute valuations, the subargument is slightly different:

VALUE3

$<a_1; e_1>$ manifests $<v_0, ..., v_m>$,

which has multi-attribute utility u_1;

so defeasibly

$<a_1; e_1>$ has utility u_1.

ANALYSIS1 is defeated by any argument that defeats any of its premises, such as a better argument that the probabilities are not what they are claimed to be, or a more specific analogy constesting the valuation of an outcome, or a multi-attribute calculation with a superset of attributes; it can be attacked by interfering with, instead of defeating, any of those same premises: e.g., by using a non-subset of attributes in the multi-attribute calculation, resulting in a different utility valuation; also, by a competing but not more specific analogy; also, by a competing but not narrower reference class for probability.

ANALYSIS1 can also be defeated by considering a larger collection of acts; likewise, interference can result, from an analysis based on a different collection of acts.

More interestingly, utility arguments can be defeated by a more detailed analysis of contingencies:

SUBANALYSIS2
In respect d_1, $<a_1; e_1>$ is like l_1, which has utility u_1;

---

so defeasibly
$<a_1; e_1>$ has utility u_1;

is defeated by

SUBANALYSIS3
In respect $d_{1,1}$, $<a1; e1; a_{1,1}>$ is like $l_{1,1}$, which has utility $u_{1,1}$;
In respect $d_{1,2}$, $<a1; e1; a_{1,2}>$ is like $l_{1,2}$, which has utility $u_{1,2}$;
$u_{1,1}$ is maximum among $\{u_{1,1}, u_{1,2}\}$

---

so defeasibly
$<a_1; e_1>$ has utility $u_{1,1}$;

and apparently by

SUBANALYSIS4
In respect $d_{1,1}$, $<a1; e1; e_{1,1}>$ is like $l_{1,1}$, which has utility $u_{1,1}$;
In respect $d_{1,2}$, $<a1; e1; e_{1,2}>$ is like $l_{1,2}$, which has utility $u_{1,2}$;
defeasibly, prob($e_{1,1}$) is $k_{1,1}$;
defeasibly, prob($e_{1,2}$) is $k_{1,2}$;

---

so defeasibly,
$<a_1; e_1>$ has utility $k_{1,1}(u_{1,1})+k_{1,2}(u_{1,2})$;

SUBANALYSIS4 might not defeat SUBANALYSIS2, however, if the analogies on which the finer outcomes' utilities are based ($d_{1,1}$ and $d_{1,2}$) are less specific than the analogies on which the coarser outcomes' utilites are based (d_1): $<a_1; e_1; e_{1,1}>$ may be like $l_{1,1}$ in SUBANALYSIS4 only superficially, while the likeness of $<a_1; e_1>$ to l_1 could be complete. Guidance here, precisely stating conditions for defeat, requires a better understanding of why deeper analsyes are supposed to be superior.

Note that the number of moments participating in the derived utility can also be disputed, as in mean-risk decision theory and the open-ended model of Allais. The premise that expected utility is appropriate for the agent could be added to each argument, supported by a best-fit subargument. This is appropriate at least in descriptive applications, and perhaps even more widely.

Behavior inconsistent with standard theories of preference can arise when decisions are reached through processes of argument. The problem arises because decisions are warranted by analyses of varying depths; uniform depth is not required of all deliberations. Act a_1 may be preferred to a_2 when arguments consider certain contingencies, but the reverse may be true when considering other contingencies. Since not all contingencies need be folded into a single comprehensive analysis, two separate deliberations could reach prima facie inconsistent preferences. If these deliberations on preferences lead to certain behaviors, such behavior violates preference axioms under uncharitable interpretations.

The same phenomenon arises in resource-bounded defeasible reasoning about probabilities.

Separate deliberations upon the probability of A and of not-A might lead to a violation of the probability axioms. This violation, would of course lead each argument to interfere with the other, so that neither warrrants its conclusion when both are constructed. But on separate occasions of deliberation, the anomaly could arise. Again, this is a violation only if interpretation makes no allowance for the revision of subjective probabilities through time.

Reconciliation of this constructive depiction of reasoning with axiomatic approaches to preference and subjective probability is not of high priority. The disparities are not so different from the non-monotonicity of conclusions over time, which is not classically logical, and which was declared at the outset to be a distinguishing feature of the present approach.

In the context of decision, it is more apt to think of Savage's problem of small worlds. The standard approach requires a fixed framing of the problem: a single description of all relevant features of the world. Savage pronounced that there was no harm in taking the modeled world to be as large as possible, but he expressed misgivings decades later. Defeasible reasoning provides one way for analyses based on larger worlds to supplant analyses based on smaller worlds.

3. CONCLUSION.

The strategy here has been to identify derivations that depend on partial information about the world; to consider that information might accumulate through computation, not just through interaction and observation; then to assert the meaningfulness of conclusions based on reasoning from the partial information.

If a derivation constructs rational belief for a reasoner who lacks information due to inexperience, then the derivation should work for an agent who lacks information due to computational limitation. Total evidence could be required for ideal reasoners and usually is. But total world experience could have been required for super-ideal reasoners. If total evidence suffices for those with limited interaction, then total accessible evidence should suffice for those with limited computation.

By assumption, reasoning based on greater information is superior. If computation is inference, and inference provides information, then reasoning based on superior computation is superior reasoning. This is not an argument for the superiority of reasoning based on superior computation, but it explains the intuition.

For statistical reasoning, the reference class is successively refined. For analogical reasoning, the basis of similarity is improved. For theory-formation, the class of candidate theories compared for quality of fit is extended. For decision under risk, the detail in the analysis of scenarios is increased. One can imagine that all reasoning can be portrayed similarly, except for the most mundane forms of non-ampliative.

When reasoners have permission to reach conclusions on less than the totality of evidence, there may be abandon. To compensate, reasoners recognize conditions under which their arguments could be defeated. This paper states many of those conditions explicitly. For each form of inference, defeasibility is based on superior computation. Restatement of these conditions in formal theories of argument may be possible, though such a reduction would obscure the main point.

Receptiveness to defeating counterargument is still not enough: the construction of arguments must proceed according to a schedule of the right kind. When computation is dominated by search for arguments, and search is goal-directed and serial, as in typical automated reasoning programs, the selection of goals for search does not favor one conclusion or another. The protocol for dispute must be fair (there would be a willingness to exchange position ex ante, or a willingness to universalize the protocol) and efficient (opportunities for counterargument are available when possible, and arguments are relevant).

Defeasibility does not merely affect the style in which theories of reasoning are formalized; it does not merely guide revisions of derived corpora of beliefs. Defeasibility can refer to computation. It permits normative theorizing about reasoners who have the frailties and failures, the bounded resources, that humans and machines possess alike.

4. REFERENCES

Alchourron C., Gaerdenfors, P., and Makinson, D. "On the logic of theory change," *Journal of Symbolic Logic 50*, 1985.

Allais, M. "The foundations of a positive theory of choice (1952)," in M. Allais and O. Hage, eds., *Expected Utility and The Allais Paradox*, Reidel, 1979.

Belzer, M. "Reasoning with defeasible principles," *Synthese 66*, 1986.

Boutilier, C. "Normative, subjunctive, and autoepistemic defaults: adopting the Ramsey Test," in *Principles of Knowledge Representation and Reasoning*, B. Nebel, et al., eds., Morgan Kaufman, 1992.

Carnap, R. "On inductive logic," *Philosophy of Science 12*, 1945.

Chisholm, R. *Theory of Knowledge*, Prentice-Hall, 1966.

Churchman, C. *Design of Inquiring Systems*, Basic Books, 1971.

Cross, C. "Belief revision, non-monotonic reasoning, and the Ramsey Test," in *Knowledge Representation and Reasoning*, H. Kyburg, et al., eds., Kluwer, 1990.

Davies, T. and Russell, S. "A logical approach to reasoning by analogy," *Proc. IJCAI*, 1987.

Delgrande, J. "An approach to default reasoning based on a first-order conditional logic," *Artificial Intelligence 36*, 1988.

Doyle, J. "Constructive belief and rational representation," *Computational Intelligence 5*, 1989.

Elgot-Drapkin, J., Perlis, D., and Miller, M. "Memory, reason, and time," in *Philosophy and AI*, R. Cummins and J. Pollock, eds. MIT, 1991.

Etherington, D. and Reiter, R. "On inheritance hierarchies with exceptions," *Proc. AAAI*, 1983.

Gabbay, D. "Theoretical foundations for non-monotonic reasoning in expert systems," in *Logics and Models of Concurrent Systems*, K. Apt, ed., Springer, 1985.

Gaerdenfors, P. *Knowledge in Flux*, MIT 1988.

Geffner, H. and Pearl, J. "Conditional entailment," *Artificial Intelligence 53*, 1992.

Ginsberg, M. "Multi-valued logcs," *Computational Intelligence 4*, 1988.

Glymour, C. and Thomason, R. "Default reasoning and the logic of theory perturbation," *Proc. Workshop on Non-Monotonic Reasoning*, New Paltz, 1984.

Goodman, N. *Fact, Fiction, and Forecast*, Harvard, 1955.

Hanson, N. "Is there a logic of scientific discovery?" in H. Feigl and G. Maxwell, *Current Issues in The Philosophy of Science*, Holt, 1961.

Harman, G. *Change in View*, MIT, 1986.

Hintikka, J. *Logic, Language Games, and Information*, Oxford, 1973.

Kemeny, J. "A logical measure function," *Journal of Symbolic Logic 18*, 1953.

Keynes, J. M. *A Treatise on Probability (1921)*, St. Martins, 1973.

Kyburg, H. *Probability and The Logic of Rational Belief*, Wesleyan U. Press, 1961.

Kyburg, H. *Logical Foundations of Statistical Inference*, Reidel, 1974.

Kyburg, H. *Epistemology and Inference*, Minnesota, 1983.

Kyburg, H. *Theory and Measurement*, Cambridge, 1984.

Lehmann, D. "What does a conditional knowledge base entail?" *Proc. of The First Intl. Conf. on Knowledge Representation and Reasoning*, R. Brachman and H. Levesque, eds., Morgan Kaufman, 1989.

Levi, I. *Enterprise of Knowledge*, MIT, 1980.

Lewis, D. *Counterfactuals*, Harvard, 1973.

Loui, R. "Defeat among arguments," *Computational Intelligence 3*, 1987.

Loui, R. "Analogical reasoning, defeasible reasoning, and the reference class," *Proc. of The First Intl. Conf. on Knowledge Representation and Reasoning*, R. Brachman and H. Levesque, eds., Morgan Kaufman, 1989.

Loui, R. "Defeasible decisions," *Uncertainty in AI 5*, M. Henrion, et al., eds., North-Holland, 1990.

Loui, R. "Ampliative inference, computation, and dialectic," in *Philosophy and AI*, R. Cummins and J. Pollock, eds. MIT, 1991.

McCarthy, J. "Circumscription," *Artificial Intelligence 13*, 1980.

McDermott, D. and Doyle, J. "Non-monotonic logic I," *Artificial Intelligence 13*, 1980.

Moore, R. "Semantical considerations on non-monotonic logic," *Artificial Intelligence 25*, 1985.

Niiniluoto, I. "Analogy and similarity in scientific reasoning," in *Analogical Reasoning*, D. Helman, ed., Kluwer, 1988.

Nute, D. "Defeasible reasoning," in *Aspects of AI*, J. Fetzer, ed., Kluwer, 1988.

Pollock, J. *Knowledge and Justification*, Princeton, 1974.

Pollock, J. "Foundations for direct inference," *Theory and Decision 17*, 1984.

Pollock, J. "Defeasible reasoning," *Cognitive Science 11*, 1987.

Pollock, J. *Nomic Probability and The foundations of Induction*, Oxford, 1990.

Pollock, J. "How to reason defeasibly," *Artificial Intelligence 57*, 1992.

Poole, D. "A logical framework for default reasoning," *Artificial Intelligence 36*, 1988.

Popper, K. *Conjectures and Refutations*, Routlege, 1963.

Prakken, H. Logical Tools For Modeling Legal Argument, doctoral dissertation, Vrije Universiteit Amsterdam, 1993.

Reichenbach, H. *A Theory of Probability*, Berkeley, 1949.

Reiter, R. "A logic for default reasoning," *Artificial Intelligence 13*, 1980

Rescher, N. *Dialectics*, SUNY Buffalo, 1977.

Russell, S. Analogical and Inductive Reasoning, doctoral dissertation, Stanford University, 1987.

Sandewall, E. "Non-monotonic inference rules for multiple inheritance with exceptions," *Proc. IEEE 74*, 1986.

Sartor, G. *Artificial Intelligence in Law*, Tano, 1993.

Savage, L. *Foundations of Statistics*, 1954.

Savage, L. "Difficulties in the theory of personal probability," *Philosophy of Science 34*, 1967.

Simari, G. and Loui, R. "A mathematical treatment of defeasible reasoning and its implementation," *Artificial Intelligence 53*, 1992.

Simon, H. *Models of Bounded Rationality v. II*, MIT, 1982.

Stalnaker, R. *Inquiry*, MIT, 1985.

Swain, M., ed. *Induction, Acceptance, and Rational Belief*, Reidel, 1970.

Thagard, P. *Computational Philosophy of Science*, MIT, 1988.

Toulmin, S. *Uses of Argument*, Cambridge, 1958.

Touretzky, D. *The Mathematics of Inheritance Systems*, Pitman, 1986.

Touretzky, D., Horty, J., and Thomason, R. "A clash of intuitions," *Proc. IJCAI*, 1987.

Vreeswijk, G. "Studies in defeasible argumentation," doctoral dissertation, Vrije Universiteit Amsterdam, 1993.

Weitzenfeld, J. "Valid reasoning by analogy," *Philosophy of Science 51*, 1984.