A Theory of Defeasible Reasoning

John L. Pollock
Department of Philosophy, University of Arizona, Tucson, AZ 85721

Reasoning can lead not only to the adoption of beliefs, but also to the retraction of beliefs. In philosophy, this is described by saying that reasoning is *defeasible*. My ultimate objective is the construction of a general theory of reasoning and its implementation in an automated reasoner capable of both deductive and defeasible reasoning. The resulting system is named "OSCAR." This article addresses some of the theoretical underpinnings of OSCAR. This article extends my earlier theory in two directions. First, it addresses the question of what the criteria of adequacy should be for a defeasible reasoner. Second, it extends the theory to accommodate reasons of varying strengths.

Reasoning can lead not only to the adoption of beliefs, but also to the retraction of beliefs. In philosophy, this is described by saying that reasoning is *defeasible*. In AI, it is described by saying that reasoning is *nonmonotonic*. My ultimate objective is the construction of a general theory of reasoning and its implementation in an automated reasoner capable of both deductive and defeasible reasoning. The resulting system is named "OSCAR." This article addresses some of the theoretical underpinnings of OSCAR.* I presented the basic ideas behind this theory of defeasible reasoning in Ref. 2. This article extends that theory in two directions. First, it addresses the question of what the criteria of adequacy should be for a defeasible reasoner. Second, it extends the theory to accommodate reasons of varying strengths.

I. PRIMA FACIE REASONS AND DEFEATERS

The basic ideas behind the present theory of defeasible reasoning come out of my work in philosophy, where I have wielded the general framework of defeasible reasoning as a tool in the analysis of a number of epistemological problems.† Reasoning proceeds by constructing arguments, where *reasons*

*A fuller description of the current state of OSCAR is presented in Ref. 1.

†The work on defeasible reasoning in philosophy stems mainly from the publications of Roderick Chisholm and myself. See Chisholm[3-5] and Pollock.[2,6-9] See also Kyburg.[10,11]

CCC 0884-8173/91/010043-22$04.00

provide the atomic links in arguments. *Conclusive reasons* logically entail their conclusions. Defeasibility arises from the fact that not all reasons are conclusive. Those that are not are *prima facie reasons*. Prima facie reasons create a presumption in favor of their conclusion, but it is defeasible. For example, something's looking red to me provides a prima facie reason for thinking that it is red. If I have no other relevant information, this makes it reasonable for me to believe that the object is red, but if I also have some independent good reason for thinking that the object is not red, that defeats the prima facie reason.

I will take a reason to be an ordered pair $\langle \Gamma, p \rangle$, where Γ is the set of premises of the reason and p is the conclusion. The simplest kind of defeater for a prima facie reason $\langle \Gamma, p \rangle$ is a reason for denying the conclusion. If for some θ, $\varphi = {\sim}\theta$, let $\neg\varphi = \theta$, and let $\neg\varphi = {\sim}\varphi$ otherwise. Then we define:

> If $\langle \Gamma, p \rangle$ is a prima facie reason, $\langle \Lambda, q \rangle$ is a *rebutting defeater* for $\langle \Gamma, p \rangle$ if and only if $\langle \Lambda, q \rangle$ is a reason and $q = \ulcorner \neg p \urcorner$.

Prima facie reasons for which the only defeaters are rebutting defeaters would be analogous to normal defaults in default logic. Experience in using prima facie reasons in epistemology indicates that there are no such prima facie reasons. Every prima facie reason has associated defeaters that are not rebutting defeaters, and these are the most important kinds of defeaters for understanding any complicated reasoning.* Defeaters that are not rebutting defeaters attack a prima facie reason without attacking its conclusion. They accomplish this by instead attacking the connection between the premises and the conclusion. For instance, ⌜x looks red⌝ is a prima facie reason for ⌜x is red⌝. But if I know not only that x looks red but also that x is illuminated by red lights and red lights can make things look red when they are not, then it is unreasonable for me to infer that x is red. Consequently, ⌜x is illuminated by red lights and red lights can make things look red when they are not⌝ is a defeater, but it is not a reason for thinking that x is not red, so it is not a rebutting defeater. Instead, it attacks the connection between ⌜x looks red⌝ and ⌜x is red⌝, giving us a reason for doubting that x wouldn't look red unless it were red. ⌜P wouldn't be true unless Q were true⌝ is some kind of conditional, and I will symbolize it as $\ulcorner P \gg Q \urcorner$. The preceding indicates that if $\langle \Gamma, p \rangle$ is a prima facie reason, then any reason for denying $\ulcorner \Pi\Gamma \gg p \urcorner$ is a defeater.† I call these *undercutting defeaters:*

> If $\langle \Gamma, p \rangle$ is a prima facie reason, $\langle \Lambda, q \rangle$ is an undercutting defeater for $\langle \Gamma, p \rangle$ if and only if $\langle \Lambda, q \rangle$ is a reason and $q = \ulcorner {\sim}(\Pi\Gamma \gg q) \urcorner$.

A useful illustration of this framework of prima facie reasons, rebutting defeaters, and undercutting defeaters, is provided by the *Statistical Syllogism*. The following defeasible reasoning schema has been much discussed in AI:

*This is illustrated repeatedly in Refs. 9, 12, and 13.

†$\Pi\Gamma$ is the conjunction of the members of a finite set Γ of propositions.

Most *F*'s are *G*'s.
This is an *F*.
―――――――
This is a *G*.

This can be represented more precisely as the following prima facie reason:

⌈*Fc* and most *F*'s are *G*'s⌉ is a prima facie reason for ⌈*Gc*⌉.

Taking prob(G/F) to be the probability of an arbitrary F being a G, this can be regarded as a qualitative version of the following principle of statistical inference:

Statistical Syllogism
If $r > .5$ then ⌈Fc & prob(G/F) $\geq r$⌉ is a prima facie reason for ⌈*Gc*⌉, the strength of the reason being a monotonic increasing function of r.*

Much work on nonmonotonic reasoning in AI has been addressed specifically at reasoning in accordance with the qualitative version of *Statistical Syllogism.* One of the central features of this reasoning is that inferences based upon more complete information about c take precedence over inferences based upon less specific information.† To use a familiar example, suppose we know:

Tweety is a bird.
Most birds can fly.
Tweety is a penguin.
Most penguins are unable to fly.
All penguins are birds.

The intuitively correct conclusion to draw is that Tweety cannot fly. By *Statistical Syllogism* we have prima facie reasons for two conflicting conclusions, viz., that Tweety can fly, and that Tweety cannot fly, but the latter takes precedence because it is based upon more specific information (Tweety's being a penguin is more specific than Tweety's being a bird, because being a penguin entails being a bird). This is diagrammed in Figure 1. The correct inference can be captured by endorsing the following undercutting defeater:

Subproperty Defeaters
⌈*Hc*, and *being an H* entails *being an F*, and it is false that most *H*'s are *G*'s⌉ is an undercutting defeater for ⌈*Fc* and the most *F*'s are *G*'s⌉ as a prima facie reason for ⌈*Gc*⌉.

*This is discussed at much greater length in Refs. 12 and 13. Some qualifications are required to make the statement of the principle correct, but the reader is referred to the above sources for the details.

†For discussions of this, see Etherington,[14] Etherington and Reiter,[15] Horty et al.,[16] Loui,[17] Nute,[18] and Touretzky.[19,20]

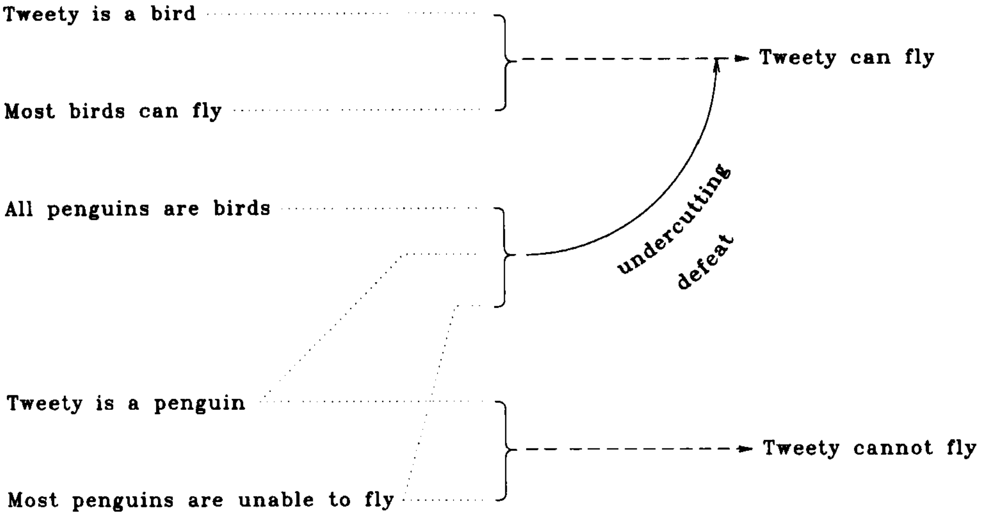

Figure 1. Subset defeators.

Given this undercutting defeater, we have prima facie reasons for both ⌈Tweety can fly⌉ and ⌈Tweety cannot fly⌉, but we also have an undercutting defeater for the former prima facie reason. Accordingly, it is removed from competition, leaving the matter undefeated.*

My experience in epistemology has convinced me that rebutting defeaters and undercutting defeaters are the only kinds of defeaters required for representing the logical structure of complicated instances of defeasible reasoning. Other proposals have been advanced by AI researchers, however. Ron Loui[17] has advocated *specificity defeaters*. Specificity defeaters are not reasons, like rebutting and undercutting defeaters. Instead, it would be better to talk about a *rule of specificity defeat*. This is a structural rule for resolving conflicts between reasons. The idea is supposed to be that if we have an argument η supporting p and an argument σ supporting $\neg p$, if the nonredundant premises of η include all of the nonredundant premises of σ, then η takes precedence over σ and defeats it. The motivation for such a rule comes from looking at subset defeaters in inferences in accordance with *Statistical Syllogism*, but I would argue that the rule of specificity defeat is mistaken in several important respects. First, it is an incorrect description of subset defeaters. It arises from looking only at cases in which we have ⌈Most H's are non-G's⌉ rather than just ⌈It is false that most H's and G's⌉. If we have only the latter, then we do not have a reason for ⌈$\sim Hc$⌉, and so the inference does not have the form described by the rule of specificity

*This account of reasoning in accordance with *Statistical Syllogism* was first proposed in Ref. 21. A somewhat similar account was proposed by Touretzky,[19] and the general idea has since been embraced by most researchers in defeasible reasoning. The full theory of reasoning in accordance with *Statistical Syllogism* requires a number of additional complexities. See Refs. 12 and 13 for details.

defeat. Second, even in cases in which we have ⌈Most H's and non-G's⌉, the inference does not have the correct form, because the rule of specificity defeat overlooks the need for the conjuncts ⌈Most F's are G's⌉ and ⌈Most H's are non-G's⌉ in the conflicting prima facie reasons. The arguments supporting ⌈Gc⌉ and ⌈$\sim Gc$⌉ must include steps supporting these statistical premises, and if they do then it is not true that the premises of the latter argument include all of the premises of the former argument. Third, *Statistical Syllogism* is just one kind of prima facie reason. I have never seen any persuasive examples illustrating that something like subset defeaters or specificity defeat operates in connection with other prima facie reasons. Finally, as I have illustrated above, cases in which something like specificity defeat seems correct are easily handled with undercutting defeaters, so no further machinery is required.

II. ARGUMENTS

Reasoning starts with premises that are input to the reasoner. (In human beings, they are provided by perception). The input premises comprise the set *input*. The system then makes inferences (some conclusive, some defeasible) from those premises using reason schemes. Reasons are combined in various patterns to form arguments. The simplest arguments are *linear* arguments. These can be viewed as finite sequences of propositions each of which is either a member of input or inferrable from previous members of the sequence in accordance with some reason schema.

It is very important to realize that not all arguments are linear. The easiest way to see this is to note that the linear arguments can only lead to conclusions that depend upon the members of input, but actual reasoning can lead to *a priori* conclusions like $(p \vee \sim p)$ or $((p \mathbin{\&} q) \supset q)$ that do not depend upon anything. What makes this possible is *suppositional reasoning*. In suppositional reasoning we "suppose" something that we have not inferred from input, draw conclusions from the supposition, and then "discharge" the supposition to obtain a related conclusion that no longer depends upon the supposition. The simplest example of such suppositional reasoning is *conditionalization*. When using conditionalization to obtain a conditional $(p \supset q)$, we suppose that antecedent, p, somehow infer the consequent q from it, and then discharge the supposition to infer $(p \supset q)$ independently of the supposition. Similarly, in *reductio ad absurdum* reasoning, to obtain $\neg p$ we may suppose p, somehow infer $\neg p$ on the basis of the supposition, and then discharge the supposition and conclude $\neg p$ independently of the supposition. Other varieties of suppositional reasoning include dilemma (reasoning by cases) and universal generalization.

In suppositional reasoning, we can no longer think of arguments as finite sequences of propositions, because each line of an argument may depend upon suppositions. We can instead think of lines of arguments as ordered pairs $\langle X, p\rangle$ where X is the set of propositions comprising what is supposed on that line. I will refer to X as the *premise set* of $\langle X, p\rangle$. Linear arguments can be viewed as arguments in which the premise sets are always empty. Discharge rules are rules that manipulate premise sets. For instance, conditionalization could be formulated as follows:

From $\langle X\cup\{p\}, q\rangle$, infer $\langle X, (p \supset q)\rangle$.

Rules of inference are really rules for the construction of arguments, so conditionalization could be stated more precisely as follows:

> If σ is an argument and some line of σ is $\langle X\cup\{p\}, q\rangle$, then $\sigma\hat{}\langle X, (p \supset q)\rangle$ (the result of appending $\langle X, (p \supset q)\rangle$ to the end of σ) is also an argument.

Other rules for argument formation will include the following:

Input
If $p\in input$ and σ is an argument, then for any X, $\sigma\hat{}\langle X, p\rangle$ is an argument.

Supposition
If σ is an argument and X is any finite set of propositions, then if $p\in X$, $\sigma\hat{}\langle X, p\rangle$ is an argument.

Reason
If σ is an argument, $\langle X, p_1\rangle, \ldots, \langle X, p_n\rangle$ are members of σ, and $\{p_1, \ldots, p_n\}$ is a reason (either conclusive or prima facie) for q, and for each i, $X_i \subseteq X$, then $\sigma\hat{}\langle X, q\rangle$ is an argument.

Dilemma
If σ is an argument containing $\langle X, (p \vee q)\rangle$, $\langle X\cup\{p\}, r\rangle$, and $\langle X\cup\{q\}, r\rangle$, then $\sigma\hat{}\langle X, r\rangle$ is an argument.

A distinction can be drawn between *factual* and *counterfactual* suppositional reasoning. In factual suppositional reasoning, we suppose that something *is* the case, and then reason about what else is the case. In counterfactual suppositional reasoning, we make a supposition of the form "Suppose it *were* true that P," and then reason about what *would* be the case. These two kinds of suppositional reasoning work in importantly different ways. In factual suppositional reasoning, because we are supposing that something *is* the case, we can automatically combine that supposition with anything we have already concluded to be the case. Counterfactual suppositions may, on the other hand, override earlier conclusions and require their retraction within the supposition. In the present context, I am only concerned with factual suppositional reasoning. Accordingly, we can formulate a rule of inference allowing the adoption within a supposition of any conclusion already adopted within a less inclusive supposition:

Foreign Adoptions
If σ is an argument, $\sigma_i = \langle X, p, \beta\rangle$, and $X \subseteq Y$, then $\sigma\hat{}\langle Y, p\rangle$ is an argument.

Note that *Reason* already builds in foreign adoptions, allowing us to directly infer conclusions from reasons adopted in less inclusive suppositions. *Dilemma* could be generalized similarly.

Reductio ad absurdum can only function in a purely deductive context. Deriving a contradiction within a defeasible argument defeats the defeasible steps rather than refuting the supposition. So to formulate a correct rule of *reductio*, we must begin by defining:

> η is a *subargument* of σ if and only if η is a subsequence of σ and η is an argument.
>
> σ is a *deductive argument* if and only if σ is an argument but σ does not employ the rule *Reason* in connection with any prima facie reasons.

We can then formulate two rules of *reductio ad absurdum:*

> *Reductio-1*
> If σ is an argument containing a deductive subargument η containing a line $\langle X \cup \{p\}, \neg p\rangle$, then $\sigma^\frown\langle X, \neg p\rangle$ is an argument.
>
> *Reductio-2*
> If σ is an argument containing a deductive subargument η containing lines of the form $\langle X \cup \{p\}, q\rangle$ and $\langle X \cup \{p\}, \neg q\rangle$, then $\sigma^\frown\langle X, \neg p\rangle$ is an argument.

An argument σ *supports* the proposition p *relative to* the supposition X if and only if $\langle X, p\rangle$ is a member of σ. σ *supports* p if and only if σ supports p relative to the empty supposition.

III. REASONING AND WARRANT

In designing an automated defeasible reasoner, one is faced with the difficult problem of how to evaluate the reasoning that the system performs. We want the reasoning performed by the system to be "correct," but what is the criterion for correctness? In answering this question, it is useful to distinguish between theories of reasoning and theories of warrant. Theories of reasoning are basically procedural theories. They are concerned with what a reasoner should do next when it finds itself in any particular epistemological situation. Correct reasoning can involve numerous false starts, wherein a belief is adopted, retracted, reinstated, retracted again, and so forth. At each stage of reasoning, if the reasoning is correct then a belief held on the basis of that reasoning is justified, even if subsequent reasoning will mandate its retraction. *Epistemic justification*, in this sense, is a procedural notion consisting of the correct rules for belief updating having been followed by the system up to the present time in connection with the belief being evaluated.

By contrast, *warrant* is what the system of reasoning is ultimately striving for. A proposition is warranted in a particular epistemic situation if and only if, starting from that epistemic situation, an ideal reasoner unconstrained by time or resource limitations, would ultimately be led to adopt belief in the proposition. Warranted propositions are those that would be justified "in the long run" if the system were able to do all possible relevant reasoning. A proposition can be justified without being warranted, because although the system has done everything correctly up to the present time and that has led to the adoption of the belief, there may be further reasoning waiting to be done that will mandate the retraction of the belief. Similarly, a proposition can be warranted without being justified, because although reasoning up to the present time may have failed to turn up adequate reasons for adopting the proposition, further reasoning may provide such reasons. Similarly, reasoning up to the present may mandate the adoption of defeaters which, upon further reasoning, will be retracted. So justification and warrant are two importantly different notions, although they are closely related.

Two schools of thought are represented in current work on nonmonotonic logic and defeasible reasoning. Most theories are members of the *semantical school*, according to which an adequate theory must be based upon a formal semantics.* The *procedural school*, on the other hand, proposes to analyze defeasible reasoning by giving a straightforward description of "how it works," in a procedural sense, and holds semantical questions in abeyance.† The semantical/procedural distinction is related to, but not quite the same as, the warrant/justification distinction. Procedural theories are typically about justification rather than warrant, and semantical theories are typically about warrant rather than justification. However, if one understands "semantical" in a fairly narrow sense so that it includes only model theoretic semantics, then there can be theories of warrant that are not semantical. The theory of warrant presented below will not be semantical in this narrow sense.‡

My own opinon is that the importance of model theoretic semantics is vastly overrated.§ Experience in formal logic has indicated that it is possible to construct model theoretic semantics for even the most outlandish logical theories. The mere existence of a model theoretic semantics shows nothing at all about the correctness of the theory. If a theory is already known to be correct, then the discovery of a formal semantics for it can be a useful technical tool in its investigation. But the formal semantics is not itself an argument for the correctness of the theory unless there is some independent reason for thinking that the semantics is correct. The systems of formal semantics that define various species of nonmonotonic logic do indeed have initial plausibility, but I would argue that their authors have sometimes been driven more by consider-

*See, for example, McCarthy;[22,23] McDermott and Doyle;[24] Moore;[25] and Reiter.[26,27]

†The most noteworthy representative of this is school Jon Doyle.[28] A more recent proponent is Ron Loui.[17]

‡Two other theories that are not semantical in this narrow sense are Delgrande[29] and Pearl.[30]

§I hold this opinion in general, not just for nonmonotonic logic. See Chapter Six of Ref. 31 for a discussion of the significance (or lack thereof) of model theoretic semantics for standard logical theories, including the predicate calculus and modal logic.

ations of formal elegance than by an appreciation of the subtleties required of defeasible reasoning for a fullblown epistemology. Before getting carried away with a semantical investigation, we should be sure that the theory described by the semantics is epistemologically realistic, and that requires attending to the nuts and bolts of how defeasible reasoning actually works. This quickly reveals that standard theories of nonmonotonic reasoning are unable to accommodate even some quite elementary defeasible reasoning. For instance, suppose we know that most birds can fly. Now suppose we know that there is a small animal in the next room and we are trying to identify it by the sounds it makes. We may reason that if it is a bird then it can fly. No AI system of nonmonotonic reasoning can accommodate this simple inference. Our actual reasoning is very simple. We *suppose* that the animal is a bird, infer defeasibly that it can fly, and then discharge the supposition (use conditionalization) to conclude that if it is a bird then it can fly.*

Our epistemological intuitions are about reasoning, and that suggests that the best we can do is build a system that mimics human reasoning. I think that that is overly restrictive, however. We do not necessarily want an automated reasoner to reason exactly the way human beings do. Resource differences may make it possible to construct automated reasoners that improve upon human reasoning. The sense in which they improve upon it must be that they update their beliefs in ways that are more efficient at bringing the set of their beliefs into conformance with the set of warranted propositions. This suggests that the target of analysis should be *warrant* rather than justification. Our intuitions about reasoning are equally about warrant, because given a general description of an epistemological situation replete with a description of *all* the relevant arguments, our intuitions can inform us about what beliefs ought to be adopted and what beliefs ought to be retracted. A characterization of what ought to be believed *given* all possible relevant arguments is a characterization of the set of warranted propositions. Such an account can be given fairly easily if we take as primitive the notion of one argument defeating another. Suppose we have an argument α supporting a conclusion P, and an argument β that defeats α. If these are the only relevant arguments, then P is not warranted. But now suppose we acquire a third argument γ that defeats β. This situation is diagrammed in Figure 2. The addition of γ should have the effect of reinstating α, thus making P warranted. We can capture this kind of interplay between arguments by defining:

> All arguments are *level 0 arguments*.
> An argument is a *level $n + 1$ argument* if and only if it is a level 0 argument and it is not defeated by any level n argument.

An argument is *in* at level n if and only if it is a level n argument. Otherwise it is *out*. Thus α, β, and γ are all in at level 0. γ is in at level 1, but neither α nor β is

*For a more sustained critique of nonmonotonic reasoning systems, see Chapter Nine of Ref. 12.

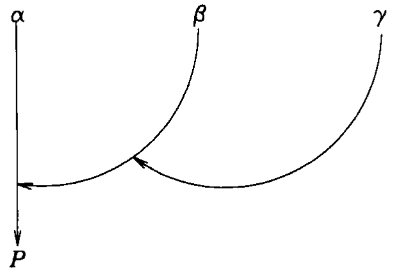

Figure 2. Interacting arguments.

in at level 1. Accordingly, α and γ are in at level 2, but β is not. And for every $n \geq 2$, α and γ are in at level n, but β is out. Let us define:

> An argument is *ultimately undefeated* if and only if there is an m such that for every $n \geq m$, the argument is in at level n.

My proposal is then that a proposition is warranted if and only if it is supported by some ultimately undefeated argument.*

Most of the rest of this article will be concerned with filling in the details in this theory of warrant. But given a theory of warrant, what are we to do with it? In an important sense, our ultimate interest in AI is in justification rather than warrant. We want to know how to build a system that reasons correctly, and that is a procedural matter. As I have just urged, the desideratum is not necessarily to build a system that replicates human reasoning in all respects, because there may be more efficient ways of doing it. However, before we can decide whether a particular procedure is a more efficient way of doing it, we have to determine what the "it" is that we want the system to do. What exactly is the connection between warrant and what we want a system of reasoning to accomplish? The simplest proposal would be that we want the system to "compute warrant." But if this is understood as requiring that the system implement an effective procedure for determining warrant, then it is an impossible desideratum. All theorems of logic are automatically warranted because the arguments suporting them are nondefeasible. This includes all theorems of the predicate calculus. However, by Church's theorem, the set of theorems of the predicate calculus is not decidable. Thus, *no* system can compute warrant in this sense. A weaker proposal would be that we want the system to generate all warranted propositions in some effective way, analogous to the manner in which a complete theorem prover generates all theorems of the predicate calculus. But this desideratum is also provably unsatisfiable. This is because, as has been observed by numerous authors,† on any theory of defeasible reasoning, the ı

*This characterization of warrant was presented in Refs. 2 and 9. A similar proposal is contained in Horty, Thomason, and Touretzsky.[16]

†I think that the first was David Israel.[32]

mate correctness of a piece of reasoning (i.e., whether the conclusion of the reasoning will survive an indefinite amount of further reasoning and hence be warranted) may turn upon something *being unprovable*, and if our resources for proof include at least the full predicate calculus, then there is no effective test for unprovability. More precisely, by Church's theorem, the set of invalid formulas of the predicate calculus is not recursively enumerable (r.e.). It follows that, for example, in default logic, a first order theory with normal defaults may have a set of theorems that is not r.e., and hence there can be no effective procedure for generating that set of theorems. The analogous conclusion applies to all theories of defeasible reasoning and to all nonmonotonic logics.

If the desideratum for an automated reasoning system is not that of computing warrant, what is it? We want the system to systematically modify its belief set so that it comes to approximate the set of warranted propositions more and more closely. We want the set of beliefs to "approach the set of warranted propositions in the limit." I propose that we understand this on analogy to the standard ε/δ definition of limits in mathematical analysis. The precise desideratum for an automated reasoner is that justification should come to approximate warrant in the following sense:

> The rules for reasoning should be such that:
> (1) if a proposition p is warranted then the system will eventually reach a point where p is adopted and stays adopted;
> (2) if p is unwarranted then the system will eventually reach a point where p is not adopted and stays unadopted.

So the task of a reasoner is not to compute warrant. It is to generate successive sets of beliefs that approximate warrant more and more closely, in the above sense. We can make this mathematically precise as follows. Define:

> A set A is *defeasibly enumerable* if and only if there is an effectively computable set function f and a recursive set A_0 such that if we define $A_{i+1} = f(A_i)$ then:
> (1) $(\forall x)$ if $x \in A$ then $(\exists n)(\forall m > n)\ x \in A_m$;
> (2) $(\forall x)$ if $x \notin A$ then $(\exists n)(\forall m > n)\ x \notin A_m$.

I will say that the pair $\langle A_0, f\rangle$ is a *d.e. approximation* of A. The intuitive difference between recursively enumerable sets and defeasibly enumerable sets is that recursively enumerable sets can be "systematically approximated from below," while defeasibly enumerable sets can be systematically approximated from above and below simultaneously. More precisely, if A is r.e., then there is an effectively computable sequence of sets A_i such that

> (1) $(\forall x)$ if $x \in A$ then $(\exists n)(\forall m > n)\ x \in A_m$;
> (2) $(\forall x)$ if $x \notin A$ then $(\forall m)\ x \notin A_m$.

The sets A_i approximate A from below in the sense that they are all subsets of A and they grow monotonically, approximating A in the limit. If A is defeasibly enumerable, however, the sets A_i need not be subsets of A. They may only approximate A from above and below simultaneously, in the sense that they may contain elements not contained in A. Every such element must eventually be taken out of the A_i's, but there need not be any point at which they have *all* been removed. The process of d.e. approximation can be pictured by thinking of A as a spherical region of space and the A_i as representing successive stages of a reverberating elastic ball whose center coincides with the center of A. As the reverberations dampen out, the outer surface of the ball will come to approximate that of the spherical surface more and more closely, but there will never be a point at which the ball is contained entirely within the spherical surface.

The reverberating sphere metaphor can be used to give a precise mathematical characterization of the difference between A being r.e. (approximation from below) and d.e. (approximation from above and below). If A is r.e. then

$$A = \bigcup_{n\in\omega} A_n.$$

On the other hand, if A is d.e. then what we have is:

$$A = \bigcap_{n\in\omega} \bigcup_{m\geq n} A_m = \bigcup_{n\in\omega} \bigcap_{m\geq n} A_m.$$

To illustrate that defeasibly enumerable sets need not be r.e., choose a pair of symbols "□" and "◇" and define:

$$A = \{\Box\varphi | \varphi \text{ is a valid formula of the predicate calculus}\}$$
$$\cup\{\Diamond\varphi | \sim\varphi \text{ is an invalid formula of the predicate calculus}\}.$$

By Church's theorem, A is not r.e., but it can be shown to be defeasibly enumerable as follows. By the completeness theorem, there is an enumeration $\varphi_i(i\in\omega)$ of the valid formulas of the predicate calculus. If for some θ, $\varphi = \sim\theta$, let $\neg\varphi = \theta$, and let $\neg\varphi = \sim\varphi$ otherwise. Then define:

$$A_0 = \{\Diamond\varphi | \varphi \text{ is a formula}\}$$
$$A_{i+1} = f(A_i) = (A_i \cup \{\Box\varphi_i\}) - \{\Diamond\neg\varphi_i\}.$$

Then despite the fact that A is not r.e., $\langle A_0, f\rangle$ is a d.e. approximation of A. The d.e. approximation in this example has a particularly simple form in which once an element is added to or removed from some A_i (for $i > 0$), its status never changes. The general form of a d.e. approximation allows items to be repeatedly added and deleted. Notice that human reasoning works in the latter way. As our reasoning develops, beliefs may be repeatedly retracted and reinstated.

The proposal regarding reasoning and warrant is that the set of warranted propositions is defeasibly enumerable, and the rules for reasoning are rules for successively approximating warrant in this way, i.e., they are rules for constructing a d.e. approximation. More accurately, we can think of a reasoner as a belief updater that operates repeatedly on a set of beliefs to generate a new set of beliefs. The reasoner starts with the set *input*, and each cycle of the reasoner constitutes the application of an effective set function *f* to the previous set of beliefs. I will say that the reasoner *provides a d.e. approximation to warrant* if and only if ⟨*input*, *f*⟩ is a d.e. approximation to the set of propositions that are warranted given that set of inputs. This is the criterion of correctness for a reasoner. This characterization of reasoning enables us to exploit differences between people and machines with regard to resource limitations. A rational machine need not reason in precisely the same way people do, because it may be better at some tasks. Thus the construction of an AI theory of reasoning divides into two parts. We must begin with a theory of warrant, and then given that characterization of warrant we can look for rules for belief updating that provide a d.e. approximation to warrant.

A consequence of this proposal is that we cannot expect an automated reasoning system to ever stop reasoning. It can inform us that "so far" a certain conclusion is justified, but it may have to continue forever in a possibly fruitless search for defeating arguments. This, of course, is just the way people work. It is not as crippling as it sounds, because once a conclusion becomes justified, it is reasonable to accept it provisionally and act upon it. This is what defeasible reasoning is all about. A common misconception in AI theories of nonmonotonic reasoning has been that before it is reasonable to act on the conclusion of some defeasible reasoning, it must be *established* that there are no true defeaters. As that is generally an impossible requirement, it has seemed mysterious how nonmonotonic reasoning can possibly function in a finite agent.

IV. UNIFORM REASONS

I have characterized warrant in terms of levels of arguments, where the latter notion is defined in terms of one argument defeating another. To complete the theory of warrant, we must characterize when arguments defeat one another. A general treatment of defeat among arguments involves addressing a complex issue that has rarely been addressed in either philosophy or AI. Reasons differ in strength. Some reasons are better than others. If we have a reason for p and a reason for $\neg$p, but the latter is significantly stronger than the former, then it wins the competition and we should believe $\neg$p. Thus a general theory of reasoning requires us to talk about the strengths of reasons and how those strengths affect interactions between reasons. One of the main objectives of this article is to address that issue, but it is expedient to begin by giving an account of defeat among arguments that ignores relative strengths. Let us adopt the simplifying assumption that all reasons are of the same strength.

An argument σ defeats an argument η by supporting a defeater for some line of η. Recall, however, that in suppositional reasoning, different lines of an argument may depend upon different suppositions. If a prima facie reason ⟨Γ, p⟩ is used in η, the presence in σ of a defeater for this reason does not guarantee that σ defeats η unless the defeater is supported in σ relative to the suppositions tion made in the context in which the prima facie reason is employed in η. For instance, the defeater might be introduced into σ as a mere supposition, where that supposition is not included in the supposition set of the line of η on which $\langle \Gamma, p \rangle$ is used. In general, the occurrence in σ of a defeater on a line whose supposition set is Y only defeats a use of $\langle \Gamma, p \rangle$ in η on a line whose supposition set is X if $Y \subseteq X$. Accordingly, we can define:

An argument σ *rebuts* an argument η iff:
(1) η contains a line $\langle X, q \rangle$ obtained by the rule *Reason* from some earlier lines $\langle X, p_1 \rangle, \ldots, \langle X, p_n \rangle$ where $\{p_1, \ldots, p_n\}$ is a prima facie reason for q; and
(2) σ contains a line $\langle Y, \neg q, \beta X$ where $Y \subseteq X$.
An argument σ *undercuts* an argument η iff:
(1) η contains a line $\langle X, q \rangle$ obtained by the rule *Reason* from some earlier lines $\langle X, p_1 \rangle, \ldots, \langle X, p_n \rangle$ where $\{p_1, \ldots, p_n\}$ is a prima facie reason for q; and
(2) σ contains a line $\langle Y, \sim ((p_1 \& \ldots \& p_n) \gg q), \beta \rangle$ where $Y \subseteq X$.

Then it seems reasonable to propose:

An argument σ *defeats* an argument η iff σ either rebuts or undercuts η.

Note that this simple analysis only works subject to the assumption that all reasons are of the same strength. Otherwise σ might contain a weaker reason for $\neg q$ than η contains for q, in which case it might fail to defeat η.

V. TAKING STRENGTH SERIOUSLY

Now consider what happens when we relax the simplifying assumption that all reasons are of the same strength.

A. Measuring Strength

If we are to take strength seriously, we must have some way of measuring it. One way of measuring strength is to compare reasons with a set of standard equally good reasons that have numerical values associated with them in some determinant way. I propose to do that by taking the set of standard reasons to consist of instances of *Statistical Syllogism*. For any p, F, and G, $\lceil Fc \,\&\, \mathrm{prob}(G/F) = r \,\&\, (p \equiv \sim Gc)\rceil$ provides a basis for believing $\neg p$, the strength of this basis being a function of r. Thus if X is a prima facie reason for p, we can measure the strength of this prima facie reason in terms of that value of r such that the conflicting reason $\lceil Fc \,\&\, \mathrm{prob}(G/F) = r \,\&\, (p \equiv \sim Gc)\rceil$ exactly counteracts it. We could take r itself to be the measure of the strength of the reason, but it turns out that a somewhat more convenient measure is $\log(.5/1 - r)$. For $.5 \leq r \leq 1$, this is a monotonic increasing function of r. This proposal has the convenient consequence that the strength of an instance of *Statistical Syllo-*

gism in which $r = .5$ is 0, and as we will see below, it makes plausible a certain principle for evaluating the outcome of conflicting reasons. So my proposal is:

> If X is a prima facie reason for p, the strength of this reason is $\log(.5/1 - r)$ where r is that real number such that an argument for $\neg p$ based upon the supposition $[Fc \;\&\; \mathrm{prob}(G/F) = r \;\&\; (p \equiv {\sim}Gc)]$ and employing *Statistical Syllogism* exactly counteracts the argument for p based upon the supposition X.

For instance, if we decide that a prima facie reason of minimal acceptable strength corresponds to an instance of *Statistical Syllogism* in which $r = .95$, then it follows that prima facie reasons must have a strength ≥ 1. Let (reason-strength $X\ p$) be the strength of a prima facie reason $\langle X, p\rangle$.

Conclusive reasons logically guarantee the truth of their conclusions given the truth of the premises, so there can be no accompanying attenuation in strength of justification. We can capture this by taking them to have infinite strength.

B. The Weakest Link Principle

Given a measure of the strengths of reasons, what are we to do with it? Strengths are important in deciding whether a reason is strong enough to justify a belief. The simplest way to deal with this is to say that a reason that is too weak is not a reason at all. This has the consequence that, for most purposes, the system need not worry about the strengths of reasons. It can just assume that all reasons are sufficiently strong to justify their conclusions.

However, there is a residual problem. Although a reason is guaranteed to be sufficiently strong to justify its conclusion in a one step argument, how do we determine the justification of the conclusion of a complex argument that involves a number of inferences? It is often supposed that each inference attenuates the strength of the conclusion, and so, although each reason by itself is sufficiently strong, the degree of justification of the ultimate conclusion may be too weak to justify adoption.

This is often coupled with a probabilistic model of reasoning according to which reasons make their conclusions probable to varying degrees, and the ultimate conclusion is warranted only if it is made sufficiently probable by the cumulative reasoning. This has the untoward consequence that even purely deductive reasoning from a set of warranted premises need not guarantee the warrant of the conclusion. Principles like *adjunction* and *modus ponens*, which deduce a conclusion from a set of more than one premise, cease to be correct principles of reasoning. This is because, by the probability calculus, $(P \;\&\; Q)$ need not be as probable as either P or Q, and Q need not be as probable as either P or $(P \supset Q)$. All we can be sure of is that $\mathrm{prob}(P \;\&\; Q) \geq \mathrm{prob}(P) + \mathrm{prob}(Q) - 1$, and $\mathrm{prob}\ (Q) \geq \mathrm{prob}(P) + \mathrm{prob}(P \supset Q) - 1$.

As a description of human reasoning, this seems clearly wrong. Once one has arrived at a set of conclusions, one does not hesitate to draw further deductive conclusions from them. The probabilistic model is just mistaken as a description of human reasoning. Furthermore, there is a good computational reason why this must be the case. Gilbert Harman[33] points out that the probability of a conjunction is not a function of the probability of its conjuncts. To compute prob(P & Q) we must know the probability of one of the conjuncts, e.g., prob(P), and we must also know the conditional probability prob(Q/P). More generally, to reason in accordance with the probability calculus about a set of propositions Γ, for each P in Γ and for each subset Λ of Γ, we must know the conditional probability prob($P/\Pi\Lambda$). These are independent probabilities, and cannot be computed from anything simpler. If Γ has n members, there will be 2^n such conditional probabilities. To illustrate that this is an impossible requirement for any real cognizer, Harman points out that if Γ consists of just 300 propositions (a very small number for any realistic system), this will require the system to store 10^{90} conditional probabilities. It can be better appreciated what a large number this is by noting that it has been estimated that there are just 10^{78} elementary particles in the entire universe. Obviously, no real system could work this way, including human beings.*

The moral to this story is that the degree of justification of the conclusion of a deductive argument cannot be such a complex function of the strengths of the premises. Computing the degree of justification of the conclusion must be computationally relatively simple. My suggestion is that this is actually done in terms of the *Weakest Link Principle*, according to which a deductive argument is as good as its weakest link. More precisely:

> The degree of justification of the conclusion of a deductive argument is the minimum of the degrees of justification of its premises.

This principle applies only to deductive arguments, but we can use it to obtain an analogous principle for defeasible arguments. If P is a prima facie reason for Q, then we can use conditionalization to construct a simple defeasible argument for the conclusion ($P \supset Q$), and this argument turns upon no premises:

Suppose P

Then (defeasibly) Q.

As this argument has no premises, the degree of justification of its conclusion should be a function of nothing but the strength of the prima facie reason. The next thing to notice is that any defeasible argument can be reformulated so that prima facie reasons are only used in subarguments of this form, and then all subsequent steps of reasoning are deductive. The conclusion of the defeasible argument is thus a deductive consequence of members of *input* together with a

*For a more extensive discussion of probabilism in epistemiology, see Pollock.[9]

number of conditionals justified in this way. By the weakest link principle for deductive arguments, the degree of justification of the conclusion should then be the minimum of (1) the degrees of justification of the members of input used in the argument and (2) the strengths of the prima facie reasons.

There are two ways of assigning degrees of justification to the members of input. One natural proposal would take input to consist of propositions like "That looks red to me," from which one can infer defeasibly "That is red." Because something can look more or less clearly red, and that can affect the justification of the conclusion, this will require us to assign differing degrees of justification to "That looks red to me," these degrees depending upon how clearly it looks red. But a simpler alternative, which I will adopt here, is to build the degree of justification into the member of input itself, so that it does not include propositions like "That looks red to me," but rather proposition like "That looks clearly red to me," "That looks vaguely red to me," etc. Then the differing degrees of justification attaching to the conclusion "That is red" can be regarded as resulting from the fact that the different input propositions provide prima facie reasons of differing strengths for this conclusion. We can thus ignore degrees of justification for members of input in computing the degree of justification for the conclusion of a defeasible argument, and identify the latter degree of justification with the minimum of the strengths of the prima facie reasons employed in the argument.* This is *The Weakest Link Principle for Defeasible Arguments:*

> The degree of justification of the conclusion of a defeasible argument is the minimum of the strengths of the prima facie reasons employed in it.

The problem of computing degrees of justification is thus computationally simple. Sometimes, it will be convenient to talk about the *strength of the argument* as being the degree of justification of its conclusion.

Where $\langle X, p \rangle$ is a prima facie reason and the members of X are justified to varying degrees, the strength of the argument that results from combining the prima facie reason $\langle X, p \rangle$ with the arguments supporting the members of X will be the minimum of (reason-strength $X\ p$) and the degrees of justification of the members of X. It will be convenient to define:

> (strength $X\ p$) = minimum{δ, (reason-strength $X\ p$)} where δ is the minimum of the degrees of justification of the members of X.

Thus (reason-strength $X\ p$) is the strength of the reason in isolation, whereas (strength $X\ p$) is the degree of justification p acquires from X in the actual epistemiological situation.

*Everything said about the members of *input* here can equally be said about propositions that are introduced by supposition in suppositional reasoning. For technical reasons, I will assign them infinite degrees of justification.

C. Comparing Competing Reasons

The most important role of the strengths of reasons lies in deciding what to believe when one has conflicting arguments for p and $\neg p$. Then one must consider which argument is better, and if it is *enough* better, one should reason in accordance with it and draw the corresponding conclusion. The question is then, "What constitutes being enough better?" My proposal is that this can be captured by an analogue of *The Likelihood Principle* from statistics. I argued in Pollock[12] that this principle is best understood in terms of the following variant of the statistical syllogism:

> *Inverse Statistical Syllogism*
> If $r > .5$, then $\ulcorner{\sim}Gc$ & $\text{prob}(G/F) \geq r\urcorner$ is a prima facie reason for $\ulcorner{\sim}Fc\urcorner$, the strength of the reason being the same function of r as in *Statistical Syllogism.*

Inverse Statistical Syllogism stands to *Statistical Syllogism* as *modus tollens* stands to *modus ponens.*

In statistical reasoning, it is very common to have an argument for p based upon an instance of *Inverse Statistical Syllogism* in which the probability is r, and an argument for $\neg p$ based upon an instance of *Inverse Statistical Syllogism* in which the probability is s. The *likelihood ratio* of p is then $(1 - r)/(1 - s)$. *The Likelihood Principle* tells us that the on-balance justification for p is measured by the likelihood ratio. This is not an entirely uncontroversial principle, but it is plausible and has played a role in the foundations of statistical inference.* Assuming the Likelihood Principle, it is reasonable to believe p in this case if and only if $(1 - r)/(1 - s)$ is sufficiently small.† In other words, there must be some ξ such that it is reasonable to believe p if and only if:

$$(1 - r)/(1 - s) < \xi.$$

Equivalently:

$$\log(.5/1 - r) - \log(.5/1 - s) > -\log(\xi).$$

In statistical reasoning, ξ is typically taken to be in the vicinity of .1 or .01, in which case $-\log(\xi)$ is in the vicinity of 1 or 2. Given our measure of the strengths of reasons, the Likelihood Principle can then be formulated equivalently as follows:

> There is a δ_0 such that, given an argument for p based upon an instance of *Inverse Statistical Syllogism* of strength η and an argument for ¬p based

*See the discussion in my Ref. 12.

†This is intuitively backwards. It would be more natural to take the likelihood ratio to be $(1 - s)/(1 - r)$, but that is not the way it has traditionally been done in statistics.

upon an instance of *Inverse Statistical Syllogism* of strength ν, adoption of p is reasonable if and only if $\eta - \nu > \delta_0$.

Reasonable values of δ_0 will be around 1 or 2.

I propose to generalize this principle and apply it to all cases of rebutting defeat:

> *The Principle of Rebuttal Resolution*
> There is a δ_0 such that, given an argument for p of strength η and an argument for $\neg$p of strength ν, adoption of p is reasonable if and only if $\eta - \nu > \delta_0$.

D. The Accrual of Reasons

If we have two independent reasons for a conclusion, does that make the conclusion more justified than if we had just one? It is natural to suppose that it does, but upon closer inspection that becomes unclear. Cases that seem initially to illustrate such accrual of justification seem upon reflection to be better construed as cases of having a single reason that subsumes the two separate reasons. For instance, if Jones tells me that the president of Slobovia has been assasinated, that gives me a reason for believing it; and if Smith tells me that the president of Slobovia has been assasinated, that also gives me a reason for believing it. Surely, if they both tell me the same thing, that gives me a better reason for believing it. However, there are considerations indicating that my reason in the latter case is not simply the conjunction of the two reasons I have in the former cases. Reasoning based upon testimony is a straightforward instance of *Statistical Syllogism*. We know that people tend to tell the truth, and so when someone tells us something, that gives us a prima facie reason for believing it. This turns upon the following probability being reasonably high:

$$\text{prob}(p \text{ is true}/S \text{ asserts } p). \qquad (1)$$

When we have the concurring testimony of two people, our degree of justification is not somehow computed by applying a predetermined function to the latter probability. Instead, it is based upon the quite distinct probability

$$\text{prob}(p \text{ is true}/S_1 \text{ asserts } p \text{ and } S_2 \text{ asserts } p \text{ and } S_1 \neq S_2). \qquad (2)$$

The relationship between (1) and (2) depends upon contingent facts about the linguistic community. We might have one community in which speakers tend to make assertions completely independently of one another, in which case (2) > (1); and we might have another community in which speakers tend to confirm each other's statements only when they are fabrications, in which case (2) < (1). Clearly, our degree of justification for believing p will be different in the two linguistic communities. It will depend upon the value of (2), rather than being some function of (1).

All examples I have considered which seem initially to illustrate the accrual of reasons turn out in the end to have this same form. They are all cases in which we can estimate probabilities analogous to (2) and make our inferences on the basis of *Statistical Syllogism* rather than on the basis of the original reasons. Accordingly, I doubt that reasons do accrue. If we have two separate arguments for a conclusion, the degree of justification for the conclusion is simply the maximum of the strengths of the two arguments. This will be my assumption.

There is a related question. Suppose we have an argument of strength η for p, and a rebutting argument of strength η for $\neg p$. If $\eta - \nu > \delta_0$, then we can go on to draw further conclusions from p. How do we compute their degrees of justification? Specifically, are they diminished by having the conflicting argument for $\neg p$? I am inclined to think that they are not. If they were, then if we acquired a second argument for $\neg p$, it would face off against a weaker argument for p and so be better able to defeat it. But that is tantamount to taking the two arguments for $\neg p$ to result in greater justification for that conclusion, and that is just the principle of accrual. So it seems that if we are to reject the latter principle, then we should also conclude that arguments that survive rebuttal by conflicting arguments are not thereby diminished in strength.

E. Defeat Among Arguments

Now we can generalize our analysis of defeat among arguments to accommodate varying strengths for reasons. It is convenient to revise our understanding of arguments by explicitly including the degree of justification ν of a line as part of the line, taking lines to be triples $\langle X, p, \nu\rangle$. These degrees of justification are computed in accordance with the weakest link principle. The rules for argument formation (rules of inference) must then be revised in obvious ways. For example:

Conditionalization
If σ is an argument and some line of σ is $\langle X \cup \{p\}, q, \nu\rangle$, then $\sigma\hat{}\langle X, (p \supset q), \nu\rangle$ is also an argument.

Input
If $p \in input$ and σ is an argument, then for any X, $\sigma\hat{}\langle X, p, \infty\rangle$ is an argument.

Supposition
If σ is an argument and X is any finite set, then if $p \in X$, $\sigma\hat{}\langle X, p, \infty\rangle$ is an argument.

Reason
If σ is an argument, $\langle X_1, p_1, \eta_1\rangle, \ldots, \langle X_n, p_n, \eta_n\rangle$ are members of σ, and $\{p_1, \ldots, p_n\}$ is a reason of strength ν for q, and for each i, $X_i \subseteq X$, then $\sigma\hat{}\langle X, q, \text{minimum}\{\eta_1, \ldots, \eta_n, \nu\}\rangle$ is an argument.

Dilemma

If σ is an argument containing $\langle X, (p \vee q), \eta\rangle$, $\langle X\cup\{p\}, r, \nu\rangle$, and $\langle X\cup\{q\}, r, \mu\rangle$, then $\sigma\hat{}\langle X, r, \text{minimum}\{\eta, \nu, \mu\}\rangle$ is an argument.

Reductio-1

If σ is an argument containing a deductive subargument η containing a line $\langle X\cup\{p\}, \neg p, \infty\rangle$, then $\sigma\hat{}\langle X, \neg p, \infty\rangle$ is an argument.

Reductio-2

If σ is an argument containing a deductive subargument η containing lines of the form $\langle X\cup\{p\}, q, \infty\rangle$ and $\langle X\cup\{p\}, \neg q, \infty\rangle$, then $\sigma\hat{}\langle X, \neg p, \infty\rangle$ is an argument.

We are finally in a position to give a characterization of defeat among arguments that takes account of the strengths of reasons. Undercutting defeat works as before—considerations of strength are not relevant because the defeated argument does not in turn provide a source of defeat for the defeating argument. Rebutting defeat is handled in accordance with the *Principle of Rebuttal Resolution*, according to which, given an argument of strength α supporting q and an argument of strength β supporting $\neg q$, it is reasonable to adopt q if and only if $\alpha - \beta \geq \delta_0$. Equivalently, it is unreasonable to adopt q, and hence that argument is defeated, if and only if $\alpha - \beta < \delta_0$. Therefore, we can characterize defeat among arguments as follows:

An argument σ defeats an argument η if and only if:

(1) η contains a line $\langle X, q, \alpha\rangle$ obtained by the rule *Reason* from some earlier lines $\langle X_1, p_1, \alpha_1\rangle, \ldots, \rangle X_n, p_n, \alpha_n\rangle$ where $\{p_1, \ldots, p_n\}$ is a prima facie reason for q; and

(2) σ contains a line $\langle X, r, \beta\rangle$ where either:

(a) r is $\lceil \neg q \rceil$ and $\alpha - \beta < \delta_0$; or

(b) r is $\lceil {\sim}((p_1 \mathbin{\&} \ldots \mathbin{\&} p_n) \gg q)\rceil$.

VI. CONCLUSIONS

I have proposed an analysis of warrant that accommodates reasons of varying strengths, and I have made a proposal regarding the criterion of correctness for an automated defeasible reasoner. This is that it should provide a d.e. approximation to the set of warranted propositions. Constructing such a reasoner is no trivial task, however. I have designed an implementation of the theory described here that aims at satisfying this desideratum, but it is not yet adequately tested.*

*This is described in Ref. 1.

References

1. J. Pollock, *OSCAR: A General Theory of Reasoning,* Working paper, 1988.
2. J. Pollock, "Defeasible reasoning," *Cognitive Science,* **11,** 481–518 (1987).
3. R. Chisholm, *Perceiving,* Cornell University Press, Ithaca, 1957.
4. R. Chisholm, *Theory of Knowledge,* Prentice-Hall, Englewood Cliffs, NJ, 1966.
5. R. Chisholm, *Theory of Knowledge,* 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ, 1977.
6. J. Pollock, "Criteria and our knowledge of the material world," *Philosophical Review,* **76,** 28–62 (1967).
7. J. Pollock, "The structure of epistemic justification," *American Philosophical Quarterly,* Monograph series 4, 62–78 (1970).
8. J. Pollock, *Knowledge and Justification,* Princeton University Press, Princeton, NJ, 1974.
9. J. Pollock, *Contemporary Theories of Knowledge,* Rowman and Littlefield, Totowa, NJ, 1986.
10. H. Kyburg, Jr., *The Logical Foundations of Statistical Inference,* Reidel, 1974.
11. H. Kyburg, Jr., "The reference class," *Philosophy of Science 50,* 374–397 (1983).
12. J. Pollock, *Nomic Probability and the Foundations of Induction,* Oxford University Press, New York, 1990.
13. J. Pollock, *How to Build a Person,* Bradford/MIT Press, 1989.
14. D. Etherington, Formalizing non-monotonic reasoning systems," *Artificial Intelligence* **31,** 41–86 (1987).
15. D. Etherington, and R. Reiter, "On inheritance hierarchies with exceptions," *Proceedings of AAAI–83,* 1983.
16. J. Horty, R. Thomason, and D. Touretzky, "A skeptical theory of inheritance in non-monotonic semantic nets," *Proceedings of AAAI–87,* 1987.
17. R. Loui, "Defeat among arguments: A system of defeasible inference," *Computational Intelligence,* **3** (1987).
18. D. Nute, "Defeasible reasoning: A philosophical analysis in PROLOG," *Aspects of AI,* J. Fetzer, (Ed.), Reidel, 1988.
19. D. Touretzky, "Implicit orderings of defaults in inheritance systems," *Proceedings of AAAI–84,* 1984.
20. D. Touretzsky, *The Mathematics of Inheritance Systems,* Morgan-Kaufmann, Los Altos, CA, 1987.
21. J. Pollock, "Epistemiology and probability," *Synthese,* **55,** 231–252 (1983).
22. J. McCarthy, "Circumscription—A form of non-monotonic reasoning," *Artificial Intelligence,* **13,** 27–39 (1980).
23. J. McCarthy, "Applications of circumscription to formalizing common sense knowledge," *Proceedings of the Workshop on Nonmonotonic Reasoning,* 1984.
24. D. McDermott and J. Doyle, "Non-monotonic logic I," *Artificial Intelligence,* **13** (1980).
25. R. Moore, "Semantical considerations on nonmonotonic logic," *Artificial Intelligence,* **25,** 75–94 (1985).
26. R. Reiter, "On closed world data bases," In *Logic and data bases,* H. Gallaire and J. Minker (Ed.), Plenum, New York, 1978.
27. R. Reiter, "A logic for default reasoning," *Artificial Intelligence,* **13,** 81–132 (1980).
28. J. Doyle, A truth maintenance system," *Artificial Intelligence* **12,** 231–272 (1979).
29. J.P. Delgrande, An approach to default reasoning based on a first-order conditional logic: Revised report," *Artificial Intelligence* **36,** 63–90 (1988).
30. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* Morgan Kaufmann, Los Altos, CA, 1988.
31. J. Pollock, *The Foundations of Philosophical Semantics,* Princeton University Press, Princeton, NJ, 1984.
32. D. Israel, "What's wrong with non-monotonic logic?" *Proceedings of the First Annual National Conference on Artificial Intelligence,* 1980, pp. 99–101.
33. G. Harman, *Change in View,* Bradford/MIT, 1986.