1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1986	Unknown

4. TITLE AND SUBTITLE
Real Rules of Inference

5. FUNDING NUMBERS
DAAB10-86-C-0567

6. AUTHOR(S)
Ronald Prescott Loui

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Rochester
Departments of Computer Science and Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Station
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
92-TRF-0027

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Statement A; Approved for public release; unlimited distribution.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* Israel's remarks about non-monotonic logic are used as a springboard for studying some thoughts on defeasible reasoning among contemporary philosophers of science: thoughts that A.I. workers have so far neglected. Israel's criticims, that A.I. has misconceived rules of belief fixation and revision, and his constructive comment, that we should look to the philosophers of science for understanding and inspiration, are the two central themes.

14. SUBJECT TERMS
Artificial Intelligence, Data Fusion, Inference, Probability, Philosophy, Inheritance Hierachies, Default Reasoning

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Real Rules of Inference
Drafted Summer 85 Revised Winter 85 Revised Spring 86

- for AI Magazine or other popular publication
- insert quotation device (italics) on sentences
- formerly The Base of Infallible, Corrigible Knowledge
- formerly An Essay on Epistemological Foundations

Ronald Prescott Loui
Departments of Computer Science and Philosophy
University of Rochester

abstract:
Israel's remarks [Isr80] about non-monotonic logic are used as a springboard for studying some thoughts on defeasible reasoning among contemporary philosophers of science: thoughts that A.I. workers have so far neglected. Israel's criticism, that A.I. has misconceived rules of belief fixation and revision, and his constructive comment, that we should look to the philosophers of science for understanding and inspiration, are the two central themes. (Conclusions are made in the last full paragraph).

Israel's Lament.

David Israel made some immoderate statements at AAAI 1980 about the state of A.I. research on defeasible reasoning. "What's wrong with non-monotonic logic?" he asks.

> The answer, briefly, is that the motivation behind the wonderfully impressive work . . . is based on a confusion of proof-theoretic with epistemological issues. . . . A.I. researchers interested in 'the epistemological problem' should look, neither to formal semantics nor to proof-theory; but to -- of all things -- the philosophy of science and epistemology. [Isr80]

Israel was dead right about where knowledge engineers should look for guidance. Work on non-monotonic reasoning has indeed seemed unnecessarily encumbered by the semantics and proof theory of energetic logicians. Though some of Israel's complaints can be rebutted by A.I.'s sophists, I think it's worth reflecting on Israel's comments, worth understanding his complaint and his proposal.

Sadly, the most important passages read like riddles:

> The researchers in question seem to believe that logic -- deductive logic, for there is no other kind -- is centrally and crucially involved in the fixation and revision of belief. Or to put it more poignantly, they mistake so-called deductive rules of inference for real, honest-to-goodness rules of inference.

Consider that old favorite: modus (ponendo) ponens. It

> is *not* a rule that should be understood as enjoining us as follows: whenever you believe that p and believe that if p then q, then believe that q. This, after all, is one lousy policy. What if you have overwhelmingly good reasons for rejecting the belief that q? [Isr80]

What could Israel mean by the phrase "real rules of inference," which somehow excludes modus ponens? How could a belief logic fail to recognize a computable deductive consequence of premises that are already believed? Of course there are logics of belief, like Hintikka's and Levesque's, which purport to analyze "belief." Perhaps such logics limit the scope of modus ponens or substitutivity of identicals for belief. It may be that the meaning of "belief" is not violated by holding Bp, but not Bq, when B(p -> q). That doesn't mean it's not an excellent idea to hold Bq, when the "proof" of q has been discovered. A normative theory of rational belief, it seems, should plop into its knowledge base whatever can be derived from its knowledge base, and intuitively at least, it seems that the familiar deductions count as ploppable derivations.

I shall discuss some of the epistemologists' "real" rules of inference and show what's honest and good about them. It should then become clear why there are people who think modus ponens can't be used blindly, and why automated inference, including non-monotonic inference, should be tied more closely to the work of some post-Carnapian epistemologists. It will turn out that A.I. has already re-invented some real rules of inference, but with none of the sophistication or appreciation of pitfalls found among the epistemologists. Finally, I'll puzzle over the significance to A.I. of the philosophers' "epistemological problem". Who should care and why?

Some Real Rules.

Some statements in knowledge bases are supposed to be at once infallible and corrigible. They are *infallible* because they wouldn't be treated any differently, roughly speaking, if more evidence were to support them. They aren't attached to their probabilities or degrees of confirmation, for instance. For the purposes of inference and decision, all these statements are of equal pedigree. They are *corrigible* because they can have this status revoked at a later time, perhaps in the light of additional evidence. It could have been an error to have treated a statement as if it were infallible.

This kind of knowledge can't be reflected by truth, deducibility, or extreme probability in the obvious ways. A statement, once true or of unit probability is genuinely infallible, but it is also incorrigible. Truth and deducibility are monotonic notions. Likewise, probability does not permit of revision downward by conditioning, if the unconditional probability is one. Non-extreme, high probability won't work in the obvious way either; a statement with anything less than unit probability is held fallible, and the question of its corrigibility is not even meaningful.

Philosophers have been interested in accepted knowledge or *rational* belief. The modifier, rational, is intended to distinguish the concept from any descriptive or psychological concept

of belief; whether logical or not, the descriptive concept is not welcome here. Rational belief, or acceptance, is what A.I. wants for its knowledge bases. What I take to be Israel's point is that acceptability has been confused with something else.

The epistemologist Isaac Levi has the following picture. There is the corpus, K, to which only infallible statements belong. A small subset of this is the Ur-corpus, of incorrigible, infallible statements. The rest of the corpus, the bulk of it, is corrigible. K is given more structure, but I'll try to avoid those details. Then there is a set of confirmation functions, Q, each member of which measures the degree to which statements in some formal language are confirmed with respect to the total evidence in the corpus. In present A.I. work, this is the body of probabilistic knowledge. Naturally, the degree to which any statement in the corpus is confirmed is one.

I shall adopt Levi's terminology because of its generality. Most major positions are special cases of this view: A.I.'s non-monotonic reasoning and automatic theorem proving deal exclusively with the corpus. Evidential reasoning is concerned with the confirmation functions. For evidentialists, the corpus contains only the statements used as evidence for conditioning, and perhaps meta-linguistic constraints on distributions, like the constraints on which entropy-maximizers rely. If probabilistic knowledge is real-valued and precise for you, then your Q is a singleton; if you use intervals, then your Q is the largest convex set satisfying particular upper and lower bounds.

At a time, t, the agent's $\langle K(t), Q(t)\rangle$ constitutes her credal state. There are questions about credal statics. Given a partial description of $\langle K(t), Q(t)\rangle$, to what else in K(t) must the rational agent be committed? And given a partial description again, what else is known about the Q(t) which the rational agent must adopt? The answers depend on what kinds of partial descriptions one assumes. Confirmation theory tries to construct Q(t) from K(t). Acceptance rules typically try to construct K(t) from Q(t). The confirmation theory and the acceptance rules together define an inductive logic, where induction connotes something more general than it did in the last century. Inductive logic suggests what must be in K(t) given some subset. Rules of inference borrowed from familiar deductive logics have also been supposed to describe K(t) from a given subset, though some think they aren't very good at that, notably David Israel. One might suspect the logic's rules, for instance, if they disagreed with the confirmation and acceptance rules. On the other hand, one might fashion the acceptance rule so that it conforms to the logic when combined with the confirmation theory. These alternatives will be a central point of discussion. Then there are harder questions about credal dynamics. Given some partial description of $\langle K(t), Q(t)\rangle$ and $\langle K(t+1), Q(t+1)\rangle$, what else is known about the agent's $\langle K(t+1), Q(t+1)\rangle$? Minimal revision rules address questions of this kind. Conditional logics also appear to have concerns here.

Real rules of inference are the ones that answer any of these questions. The challenge is to find a set of compatible real rules of inference. It's easy to write down some clean and simple rules that answer some of the questions, while begging the others. That's often

good enough -- especially if the questions left unanswered are unimportant in applications.

For example, in the simple Bayesian model that's so popular in engineering circles, there's no acceptance rule: one just assumes that statements are added to K(t) at t+1 by observation and only by observation. That's fine if you know what evidence to treat infallibly and you neither want nor need to infer that a statement is infallible. It's fine if, like the later Carnap [Car62], you hold that there can be no rational induction from the data. But then you'd better not be caught conditionalizing on anything that's not strictly observational. Or using an empirical universal generalization as if it were fully accepted. Or using scientific theory as if it were definitional. Or discussing the fallibility of observational knowledge. As Peter Cheeseman notices [Che85], this view "insist[s] that the only propositions that can be known with certainty are tautologies -- any empirical proposition can only be known probabilistically . . .; however, this insistence forbids . . . logically reasoning . . . about the real world!" One must thus view everything but probabilistic reasoning as an "approximation" to the real thing.

Modern philosophers of science have taken a different tack and have avoided this probabilist view. They hold that there can be promotion of a statement into K, on the basis of satisfactory confirmation values in Q. It's possible to infer that a statement must belong to K. Philosophers have worked on developing the acceptance rules required to do the promotion: rules that say when the confirmation values in Q are satisfactory for admission into K. For A.I., it means that not just any old "approximation" will work. In fact, it shouldn't be called approximation at all; a program can use reasoning other than probabilistic reasoning, and be doing the real thing. It means that a statement in a knowledge base can be simultaneously infallible (accepted in K(t)) and corrigible (removed at t+1) for reasons. A statement can be granted infallibility and have infallibility revoked on the basis of defeasible inferential rules: not just on the authority or whim of some external source (or deduction therefrom).

Acceptance rules solve one of the static questions. A.I.'s quest in the name of non-monotonicity appears to be a quest for the dynamic rules. Still, the acceptance rule together with the confirmation theory trivially gives an answer to the dynamic question that's interesting. As evidence grows, which non-monotonic inferences remain valid? Just ask what can be accepted in K(t) on the basis of some "evidential" subset, what can be accepted in K(t+1) on the basis of a slightly larger "evidential" subset, and compare the two. That's the answer. For this reason, inductive logic construed widely has lots to do with rational non-monotonic reasoning. They are so closely related that it's an interesting question whether there are any non-monotonic inferences worth making that aren't licensed by induction. A.I. writers have supposed that there are, and epistemologists can't understand why they have so supposed.

An acceptance rule is exactly the solution philosophers have proposed for Israel's epistemological problem. Soon we'll have to take a close look at why knowledge engineers should care; what has motivated

the philosophers are considerations that seem irrelevant in the field. Ironically, it'll turn out that what motivates A.I. practitioners to use the philosophers' undecidable acceptance rules is the desire to compute inference effectively in evidential situations! At the moment, a quick look at acceptance will be instructive in distinguishing rules that are really quite good epistemological maxims from those that have only been confused to be so.

So what's a good acceptance rule? The obvious one is the high-probability rule,

> Acc(h) iff Prob(h | K) > 1 - e,
> for small e;

or, in terms of Q, which can be thought of as containing many probability measures,

> Acc(h) iff for all Qi in Q, Qi(h | K) > 1 - e,
> for small e.

To philosophers, the rule is associated with W. Sellars, H. Kyburg, C. Hempel, R. Chisholm, and even Jacob Bernoulli. In A.I., we find it mentioned in Quinlan [Qui83]: "Converting the probabilities to categorical form by thresholding . . ."; and in Ascher [Asc84]: "What we do is examine the probability of each one of [our default inferences] conditional upon K U K+: those whose conditional probabilities continue to exceed a will be included in the expansion, (K U K+)*." We also saw something like it in MYCIN, when positive CF's indicated belief, and in Rich [Ric83]: "If there are no CF's represented, *then there must be a threshold CF value (not normally* explicit or consistent) below which statements must simply be thrown out by the knowledge base creator, or the statements must be refined with additional premises until their CF's cross the threshold." According to Cheeseman [Che85], "A reasonable compromise is to treat propositions whose probability is close to . . . 1 as if they are known with certainty -- i.e., thresholding probability values if they are 'beyond reasonable doubt.'"

The only problem is that for any non-zero e, this acceptance rule leads to corpora that are inconsistent in a significant sense: it is possible to accept statements p1, . . ., pk, such that p1, . . ., pk |- x ~= x. If p1 is the proposition that "ticket one will not win" in a fair lottery, and if the number of tickets in the lottery exceeds 1/e, p1 is always acceptable. So is p2, that ticket two will not win, and p3, etc. Since some ticket must win, there is a serious problem with the resulting corpus.

No one in A.I. seems to have cared about this, or even noticed this, though the philosophers have worried about it for some time. It's certainly the case that if you want to follow Quinlan, Ascher, or Rich and threshold, someone had better worry about it soon. It also puts some punch into Israel's remarks, because the lottery paradox has driven some authors to avoid even deductive consistency as a desirable property of the corpus.

Robert Nozick's acceptance rule [Noz81],

Acc(h) iff Prob(K | h) >= 1 - e
and Prob(K | ~h) <= e
and Prob(h) > Prob(K | ~h),
for small e;

and I.J. Good and A. McMichael's rule [GMc84],

Acc(h) iff log Prob(K | h) - log Prob(K) + k log Prob(h)
is sufficiently high for reasonable k;

will also accept inconsistencies, no matter what non-extreme parameters are chosen. The problem seems inherent in the thresholding method.

There are two ways to bite the bullet. One way is to require that the acceptance rule accept only consistent sets of statements. The other way is to keep the high probability rule, but weaken the inference rules that apply to statements in the corpus, so that inconsistency of this kind is harmless. Recent work in A.I. on local consistency seems to be sympathetic to the latter approach. There is the third proposal to accept statements one-by-one in some temporally significant order, so that each augments the K relative to which the next will be judged, until no more can be consistently accepted (a weak analog of Etherington and Reiter's [EtR83] solution to multiple extensions). But no one seems to like it because it reeks of arbitrariness and diachronicity. Incidentally, there are workers in A.I. who should be notified that falsificationism has passed, for its arbitrariness and diachronicity.

The new acceptance rules belong to Keith Lehrer [Leh70], Jaakko Hintikka and Risto Hilpinen [HiH66], and Levi [Lev80a]. Kyburg [Kyb70] and Hilpinen [Hil68] are the best sources for details.

Lehrer's rule accepts the most probable of the various alternatives, if there is one.

Acc(h) iff
(i) for any minimally inconsistent set of statements S
that includes h, if s is in S and different from h,
then Prob(h | K) > Prob(s | K),
and
(ii) not(K |- ~h)

S is minimally inconsistent iff S, K |- x ~= x and there is no subset of S which is minimally inconsistent.

It can be shown that K is deductively closed; i.e. if K |- h then Acc(h). The rule has been criticized for ignoring the absolute probability of h, though this is easily remedied. It accepts no statements about the lottery.

Hintikka and Hilpinen accept only universal generalizations on the basis of high confirmation, and they use a threshold on the quantity of evidence in such a way that inconsistencies don't arise.

Acc(h) iff

(a) (i) h is a universal generalization,
(ii) Prob(h | K) > 1 - e,
for small e,
and
(iii) the number of statements in K of a certain observational class exceeds some critical n0.
or
(b) K |- h.

Their critics mainly charge that the language they inherited from Carnap is too restricted (though A.I. may find it quite expressive). They too accept no statements about the lottery.

Levi uses decision theory to accept statements, where decisions are based on a kind of information measure called cognitive utilities. M(h) measures the informational utility of h.

Acc(h) iff
(a) (i) h is E-admissible
and
(ii) there is no E-admissible h', h' ~= h, which is preferred to h on lexicographic considerations.
or
(b) K |- h.

h is E-admissible iff
(i) for h = h1 v h2 v . . . v hk, each hi is a potential answer to a specific question of interest,
and
(ii) for some Qj in Q, for suitable q:
h maximizes Qj(h) + qM(h);
or equivalently,
hi is a disjunct of h iff Qj(hi) > qM(hi);

In the lottery example, Levi's behavior depends on whether utilities are all equal. Usually, the informational value of each maximally consistent extension will be the same, so the disjunction "ticket 1 wins v ticket 2 wins v . . ." will be accepted.

Kyburg's way of biting the bullet is a little different. He accepts all statements based on the high probability rule, then rejects the standard deductive closure of K, by rejecting and-introduction. If s in K, s |- t, then t in K. But if s1 in K, s2 in K, s1, s2 |- t, t in K is still not guaranteed. Kyburg gets a weak deductive closure, but not the familiar strong deductive closure; similarly, he has weak deductive consistency: if s in K then not(s |- x ~= x); but doesn't satisfy the strong version, if s1 in K and s2 in K, then not(s1, s2 |- x ~= x). In the lottery example, each of the statements "ticket 1 loses," "ticket 2 loses," etc. finds its way into the corpus. So does the tautology that some ticket must win, which makes K inconsistent. Inferences based on this corpus must be warranted by each of the maximally consistent subsets of the corpus. Or even better, the inference rules are designed to deal with the inconsistencies directly;

would-be inferences over K are defeated whenever the inconsistencies make a difference. Since the acceptance rule and confirmation theory together are sufficient to augment K, based on evidential subsets of K, deductive logic needn't be used, at least not directly for the purpose of augmenting K. So inconsistency doesn't lead to absurdity.

Now we can make sense of David Israel's remarks about modus ponens. Israel equivocates, suggesting first that not(Acc(q)), with "overwhelmingly good reasons for rejecting the belief that q," then that Acc(~q), due to "your previously settled beliefs on the basis of which you were convinced that not-q" [Isr80]. Both of the credal states suggested make sense. If we use the high probability rule, we may fail to have strong deductive closure. Prob(p) may be above threshold, and Prob(p -> q) may be above threshold, but Prob(q) might be below it. So Acc(p), Acc(p -> q), but not(Acc(q)). And if local inconsistency makes any sense to you, as clearly it does to Kyburg (see also Peter Klein [Kle85], N. Rescher and R. Manor [ReM70]), then it's even cogent to Acc(~q). To be fair, Israel probably had in mind a situation wherein ~q in K(t-1); prima facie reasons (evidence normally sufficient) for ~q in K(t); prima facie reasons for p, and for p -> q in K(t); and the question was whether q belonged to K(t). Nevertheless, the weak consistency example reveals the depth of Israel's comments.

What has happened is that epistemological intuitions favor the high probability rule so strongly that the plausibility and desirability of deductive closure has been cast in doubt. In Kyburg's system, if Acc(p) and Acc(q), then frequently Acc(p & q). But the frequency goes down with the length of the conjunction. Conjunction (or adjunction) is a good epistemological maxim only when it is limited.

Note that there are as many lousy systems of deductive logic as one can dream up which have terrible epistemological properties: their inference rules could not reasonably be expected to generate acceptable statements when other accepted statements are taken as premises. The source of confidence in familiar rules must therefore be empirical: the history of human reasoning has shown entailment from accepted premises (in familiar logical systems) to be a good indication of acceptability; hence, conjunction is normally an acceptability-preserving transformation. But the history of human reasoning has also shown it to be a bad policy to accept massive, long conjunctions of arbitrarily selected, individually accepted statements. Simultaneously using both kinds of rules can lead to disagreement. We want to know whether a massive conjunction is acceptable. Deductive rules say that it is. Appeal to the inductive acceptance rule may say that it is not, especially if acceptability depends critically on probability. When they disagree, which is right? Since the premises in the deduction are based on the inductive acceptance rule, it would seem hypocritical to permit the inductive acceptance rule to be undermined. The attempt to augment the repertoire of real inference rules leads to the inconsistency. So do not augment the primary rules.

Isn't it the case that: 1. If the premises in a deductive argument are known, then they are true; 2. If deductive rules of transformation preserve truth, then the conclusion must be true; 3. If the conclusion is true, then it must be acceptable?

However, the premises in the argument are not known; they are merely accepted or believed. In order for the contents to be known, they must be correctly justified, true beliefs. Who knows whether they are true? The contents must not be knowledge, at least not in the sense that guarantees truth. The contents are merely those statements we have ascertained to be the best candidates for knowledge via truth-conducive procedures. It is a mistake to confuse the analysis of knowledge with the recipes for ascertaining whether one knows (see [Leh74; p. 48]). The semantics of knowledge and the real rules of inference live in two separate worlds. The skeptics, who doubt that the truth condition of knowledge can be satisfied, still cannot conclude that the truth condition in the semantic analysis of knowledge must be rejected. Conversely, even when one has decided that one knows and has justified one's claim, one cannot thus be certain of the claim's truth. Apparently, "knowledge base" is a misnomer.

The truth that is preserved by a logic's transformations is a semantic truth, not an epistemological one. It's an interesting question what else it is that these rules actually preserve (see [Nut84]). Few would argue that they preserve warranted assertability. They probably take supposable statements and generate other statements which must concomitantly be supposed. They clearly do not take statements of high probability and produce other statements of high probability. Because of this, there are some who think they don't preserve acceptability. And at least Israel thinks they don't preserve the property "should be in the knowledge base."

A Pragmatic View of Infallibility.

There is clearly a dispute over the logical properties of the corpus, over what should be its closure and consistency properties. This dispute must also pertain to the knowledge base, since it is just a corpus with a less philhellenic name. The logical properties will depend on the corpus' acceptance rule. But some want to discover good acceptance rules by identifying the logical properties that a good rule respects. Hence, there is discussion about target properties.

The dispute is not related to the discussions with which A.I. is familiar regarding resource limitations, decidability, or human performance. It is a dispute over the correctness of deductive closure for corpora under idealized computational ability and decidability. It is essentially the question of whether the usual truth-preserving transformations are inherently acceptability-preserving.

Of course, the logical properties of the corpus will depend on what it's being used for. The lesson is from Levi, who writes:

> Bayesians have been s[k]eptical of inductive acceptance rules. They challenge those who are friends of acceptance to identify a use for acceptance rules. Friends of acceptance disagree among themselves concerning the properties acceptance should have; and such disagreements are not likely to be resolved by an appeal to intuition. I suggest that a more promising avenue of attack is to consider the intended applications of acceptance rules. [Lev80b]

I put it this way: infallibility is needed for inference schemata that can make no sense of graded distinctions among beliefs. These schemata are often specifiable meta-linguistically, and include conditionalization, direct inference, decision rules, and the familiar deductive rules. The logical properties a corpus should have depend on the particular schema one has in mind: what is going to be done with statements accepted into the corpus?

If the intended use of the statements in the corpus is conditionalization of confirmation functions, then the corpus should be strongly consistent and might as well be deductively closed. If the intended use is as the premises in a logical analysis, with a traditional interpretation, then the corpus may have to be strongly consistent and deductively closed. If the corpus is used for cognitive modeling, to reflect psychological commitment, it had better not be closed or strongly consistent.

If the schema is direct inference, inconsistency is ok. For direct inference to the probability of a ball's redness, one needs to find the smallest reference class to which the ball is known to belong, and use the frequency of redness in that class as the probability. If there is no uniquely smallest class, so that neither subset(C1, C2) in K, nor subset(C2, C1) in K, then something else must be done. Whatever the something is, it can also be done when an inconsistency in the corpus is detected, e.g., when subset(C1, C2) in K as well as subset(C2, C1) in K, and the classes have different frequencies, hence cannot be equal. Or suppose that the frequencies of some property Z among members of C1 is the same as the frequency of Z among members of C2. According to K, C1 is distinctly not C2, and C1 U C2 is not projectible: i.e., not a reference class. Then for the event in question, x, if "x in C1" in K and "x in C2" in K, K is inconsistent. Still, the probability is just the frequency which is the same for Z's among C1 and for Z's among C2. Accepting the weaker "x in C1 U C2" instead, to preserve consistency, would not have allowed either of the frequencies to be used, since C1 U C2 isn't a reference class.

If the use of the corpus is to generate certain kinds of explanations, there is no profound implication of both p and ~p in K. Imagine the following schema for explanations. If p is in K, it can be used in an explanation (this says nothing about requiring the absence of p's contraries). Two statements explain q(a) if each can be used in an explanation, if one is a nomological generalization of the form, (x). p(x) -> q(x), and if the other is p(a). The presence of some contrary of p(a) in K: r in K and r |- ~p(a), doesn't interfere with the explanation. It may even be that such a corpus can explain both q(a) and ~q(a). Because of the nature of this schema for explanation, it does not follow that x ~ = x can be explained, or other such absurdity.

These schemata are formalized in a first order language that includes names for sentences in the corpus, and the relation "in K", with standard rules of inference, semantics, and consistent deductions from consistent premises. But that's not the point. It is the *corpus* which is not necessarily governed by deductive notions of consistency or inference.

Levi discusses more complicated stances which take the evidential subset of the corpus to be deductively closed and strongly consistent, but the corpus in general not to be [Lev80b]. Infallibility for evidential purposes is more privileged than infallibility for practical deliberation. The former is the status for statements to be used in confirmation; the latter is for statements taken as given in a decision. The acceptance rules that govern the membership of each could differ wildly, or at least have different thresholds. With EK, the evidential subset of a corpus distinguished, there is yet another logical property discussed: feeble cogency is the property that holds just in case EK is deductively closed and strongly consistent; and if ~h in EK then not(h in K). Piecemeal cogency adds the additional requirement: if h in K then not(~h in K).

Lehrer discusses equally fine distinctions between desirable and undesirable properties of acceptance rules. Let I(h, K) stand for Acc(h) when K is the corpus, or "h may be inductively inferred from the total evidence K." Then Lehrer counts

if I(h, K) and I(g, K) then I(h, K & g), and
if I(h, K) and I(g, K) then I(h & g, K)

among the desiderata, while excluding

if I(h, K) and I(g, h & K) then I(g, K), and
if I(S, K) and S |- h then I(h, K).

Narrow minded logicians tend to think either of the whole deductive pie, or of nothing at all. What I have tried to emphasize is that there is a rich set of intermediate positions (Fagin and Halpern have appreciated this in [FaH85], but perhaps only for descriptive belief logics). Leaving familiar logical rules is not intellectual suicide. If knowledge bases are to be used as repositories of theorems in a system with an idealized conception of validity, then the rules and properties of deductive logic are just fine. If knowledge bases represent the result of empirical inquiry and are used for rational deliberation, then they should conform to the rules of inductive logic, whatever your conscience may tell you about *those* rules.

Induction and Non-Monotonicity.

Nobody should be surprised that inductive logic and A.I.'s non-monotonic reasoning are related. Lehrer distinguishes two principles:

If h may be deductively inferred from K, then for any i, h may be deductively inferred from K and i.
If h may be inductively inferred from K, then for any i, h may be inductively inferred from K and i.

The first is true and the second is false. Non-monotonicity is just a defining property of inductive inference. Lehrer [Leh70] credits Carnap's "The Aim of Inductive Logic" [Car62] with the statement that additional evidence can undermine inductive inference, but cannot undermine deductive inference (even earlier was [Car50; p. 199]).

Inductive logic has its "total evidence requirement": de jure confirmation and acceptance must be relative to a specified body of evidence, and de facto confirmation and acceptance must be relative to the total body of knowledge accepted as evidence. As the evidence

changes (even if it just grows), past inferences that a sentence is acceptable are no longer relevant and are consequently defeated. Hence inferences are defeasible: i.e., they are non-monotonic, and this implies that inductively inferred statements are corrigible.

Corrigibility is anomalous only when researchers feel obligated to use concepts and formalisms designed for monotonic reasoning and incorrigible knowledge. It's all right to try, but it's not all right to talk as if there's no alternative, or as if the other problems having to do with real rules of inference are less important.

Clark Glymour and Richmond Thomason comment on the prevailing deductivist investigation and explication of non-monotonicity:

> . . . of course, a straightforward logical account with a model-theoretic definition will not serve this purpose, since monotonicity is built into such definitions. [GlT84]

Israel obviously concurs.

There are payoffs for talking induction. In inductive logic, the deference of a type's default to the default of its sub-type is reflected in the total evidence requirement. Elaine Rich [Ric83] tried to guarantee this behavior without the total evidence requirement, using a suspicious "monotonic consistency constraint." She requires the likelihood of p3 => p4 to be greater than that of p1 => p2 (here, => is her symbol for a default or conditional rule) whenever p3 |- p1 and p4 |- ~p2, because she'd like to conclude p4 on p1 and p3. She draws the conclusions of the most likely rule when rules are in conflict, irrespective of the specificity of the antecedents of those rules. So she must impose a constraint on the likelihoods to get the deference desired.

But with the total evidence requirement, the behavior is simpler than she leads us to believe. If p4 is acceptable on evidence p3, and p1 is entailed by p3, then p4 must be acceptable on the evidence {p3, p1}. Every acceptance rule in the book conforms to this logic (where each pi is a single sentence in the language, not a set of sentences). So in Rich's example, p4 must be acceptable on evidence {p3, p1}. Her less specific default, that p2 is acceptable on p1, must defer if {p3, p1} is the total evidence.

The literature on induction is rich with concepts yet to be introduced to A.I. The more complex the inferences one is willing to talk about, the more complex the ways in which those inferences can be defeated. By moving from sound inference, where there is no concept of defeat, to defeasible inference, where defeat is central, A.I. has opened the gate to a new pen of playpals.

John Pollock distinguishes "rebutting" defeaters from "undercutting" defeaters [Pol74,83]. Rebutting defeaters attack the conclusion of a would-be inference. A.I.'s only rebutting defeater is detected inconsistency. How about rebutting on the basis of the strength of the conclusion? Conclusions too weak or too strong might be unacceptable, especially in the presence of other conclusions of different strength. Undercutting defeaters attack the connection between evidence and would-be conclusion, in virtue of which the evidence was prima facie warrant for the would-be conclusion. B is an

undercutter of A for C in conditional logic when "B > ~(A > C)" for example. A.I.'s only undercutter is the explicit exception. Inductive logicians have considered defeat mechanisms based also on the relevance and specificity of antecedents and all sorts of statistical matters: sample size, sufficiency, Bayesian information, etc.

It's surprising that default logic hasn't yet had to face the tradeoff between strength and specificity defeaters. b is an exception to the inference from a to h if

$$a : Mh \mid\!\!- h$$

and

$$a \,\&\, b : M{\sim}h \mid\!\!- {\sim}h.$$

But let a & b & c be accepted. Surely

$$a \,\&\, b \,\&\, c : M(h \vee {\sim}h) \mid\!\!- h \vee {\sim}h$$

can be a default rule, too. If rules with less specific antecedents are defeated without regard to strength, then c must be an exception to the a & b : . . . rule: an exception that doesn't lead to a contradictory conclusion, but rather to a weaker, diluted one. h v ~h is concluded on a & b & c. If specificity matters only when there is inconsistency, then c is not an exception to the a & b : . . . rule, so h is concluded on a & b & c. A system with strength defeaters apparently can't have implicit exceptions.

Everyone seems to forget detachment. It is detachment that often makes Prob(C | A) the relevant assessment of the probability of C, upon learning A. Similarly, detachment is what happens in Glymour and Thomason's system for "theory perturbation" [GlT84] when, upon adding A, they add B > C to a theory that once contained A & B > C (hence, the policies for revising the knowledge base when B is added may have changed, to include B > C).

Because of detachment, the total evidence requirement doesn't destroy all of the locality that is characteristic of familiar inference rules such as modus ponens. McDermott has been worrying about these issues [McD85]. One of the desirable features of monotonic systems is that their inference rules are local. If it is the case that "if a then b", then to infer that b, one need no more than that a. In contrast, induction requires that all inferences be made with respect to the total evidence to date. So there are no eternal conditionals of the simple form; there is only the rule that one needs a body of evidence that confirms b. However, it is a simple matter to derive relations such as K U {a} confirms b. Then, if K is the total evidence at some time, it is true that to infer b, one need simply accept a. Of course, there is the proviso that a be all that is accepted, because the inferences are defeasible (e.g., it may be that K U {a, c} fails to confirm b). One can mourn the existence of such provisos. But then one is mourning the loss of monotonicity, not the loss of locality. With provisos, attention can be focused in the locality of the new evidence; the detached evidence is irrelevant after the detachment.

We saw some fine distinctions among revision strategies when we

looked at consistency and inductive rules. Here's another fine distinction. In most normative systems, an agent's initial "confirmational commitments," together with her total evidence to date, uniquely determine her present credal state. Levi calls this "confirmational tenacity," and argues that there's no reason for it [Lev80a]. In his view, agents can spuriously revise not only their knowledge and their policies of revision, but also their confirmational commitments. So the set of confirmation functions, Q, could shift for reasons other than conditionalization followed by detachment. They could change unsystematically. The issue comes up in the design of revision policies, when one wonders whether K revised by a, then ~a, then a again, should always equal the original K.

Some Intellectual History.

How is it possible A.I. should have so much work on non-monotonicity, none of which relates to the philosophers' work on defeasible and inductive inference? The obvious answer is that A.I. was attacking a different problem, though that mustn't be right, since most people working on induction can't see any difference. The better answer is that A.I. was interested in a different kind of solution. This will turn out to be a defensible position, from which one can rebuke even Israel's vehement remarks. But if one goes to the early papers, I think a strong case can be made that there were less deliberate causes. I'll trace the philosophers' and A.I.'s parted company to a Palo Alto probability phobia, to a Canadian closed world cultural precedent, and to the MIT modal-mania.

McCarthy and Hayes [McH69] can be found legitimating what may have been a widespread opinion at the time, on the use of numerical probabilities; it certainly became the standard view once they stated it: "The information necessary to assign numerical probabilities is not ordinarily available. Therefore, a formalism that required numerical probabilities would be epistemologically inadequate." These were leaders of A.I. dismissing probability in two lines. Since the literature on inductive logic is filled with the use of numerical probabilities, inductive logic too had to be dismissed without further appraisal, despite the useful concepts it had to offer.

Last year, McCarthy restated his position (and last month, Hayes reavowed his! [Hay85]):

> Why don't we use finite probabilities combined by the usual laws? That would be fine if we had the numbers, but circumscription is us[e]able when we can't get the numbers or find their use inconvenient. Note that the general probability that a bird can fly may be irrelevant, because we are interested in the facts that influence our opinion about whether a particular bird can fly in a particular situation.
>
> Moreover, the use of probabilities is normally considered to require the definition of a sample space, i.e. the space of all possibilities. Circumscription allows one to conjecture that the cases we know about are all that there are. However, when additional cases are

found, the axioms don't have to be changed. [McC85]

This is a good argument against the use of determinate probability models as a substitute for inductive reasoning in A.I. But of course, there's no reason to confine oneself to undergraduate mathematical probability theory. Once one recognizes that an inductive logic can be based on measures that are not real-valued, or on indeterminate probabilities, the "can't get the numbers" argument is enfeebled. I think the convenience argument is a tenable one, but it is certainly not the tale of epistemological inadequacy he wants to tell. From the philosopher's point of view, what is really epistemologically inadequate is determining non-inferentially which predicates may be circumscribed and which default rules are acceptable.

McCarthy seems to have problems determining the reference class for a particular bird with respect to flight. He supposes that we would want to appeal to "the general probability that a bird can fly . . .." Perhaps he supposes this because we may never have sampled from the class of birds identical to this particular bird. He reminds us that the general probability may not be relevant; relevance must be determined. Philosophers have always faced this fact squarely; it is the mathematicians who have idealized the relevance problem away. The probability of flight for this bird is determined by the information that we have that is relevant.

Reichenbach thought that the probability of the statement of an event comes directly from estimates about the narrowest class containing the event, *of those classes about which adequate estimates are accepted*. The last qualification is important, since any event will always belong to the singleton class containing only the event itself, and this singleton will always be narrowest. Informative statistics are not ordinarily known about this class. In these cases, the knowledge that the event belongs to its own singleton should not interfere with the calculation. Note that this is different from the Bayesian who is obliged to condition on all known class memberships of the event -- to wit, on all that is known -- even though the resulting, ultimately particular conditional may be poorly estimated or impossible to estimate: the Bayesian has her own ways of constructing probabilities in these situations (of which Cheeseman has told us [Che85]). Anyway, the non-Bayesian, Reichenbachian, "epistemological" conception of probability (see [Kyb74]) has the *krestomatheia* McCarthy desires, allowing that only useful knowledge should enter into the calculation of a probability. And it's not sensitive to the form and number of unobserved possibilities in the "sample space."

It may still be true that "enough information" to assign useful numerical probabilities is not available, or that in some domains, the numbers are simply unimportant. But don't poke your eyes out because it's dark and you can't see. There may be enough information in some domains to assign useful bounds on measures, and it may be good enough to know where these bounds lie. If you stick to an indeterminate numerical formalism, then when there is enough information to get narrow bounds, in McCarthy's words, "the axioms don't have to be changed." What McCarthy and Hayes should say is that A.I. has developed an alternative style of formalizing inference, with different

tradeoffs between expressiveness and convenience: one that doesn't bother with the numbers. They should not be talking of epistemological inadequacy.

From the fact that McCarthy was attacking an unworthy opponent, determinate probability representations, it's tempting to conclude that McCarthy simply doesn't see who the worthy opponents are. More likely, the inconvenience of numerical calculations was so terrible a thought that epistemological arguments had to be invented against quantitative approaches.

The strong possibility that the A.I. world was simply ignorant of what inductive logic had to offer was most poignantly manifest in Winograd's comment at a conference for applications of inductive logic:

> Inductive logic has not traditionally dealt with the problem of the acceptance of a conclusion which is not certain. In a practical situation, it is often necessary to act as if a particular conclusion were true, even though the formal rules of evidence can do no more than assign it a plausibility or demonstrate that is has not been falsified. [Win80a]

Winograd's complaint is exactly true of the ongoing A.I. work in evidential reasoning in the probabilist or Bayesian mold. But it is simply false to say of inductive logic that it does not deal with the problem of acceptance; as I have illustrated, the acceptance problem is inductive logic's central issue: indeed, its defining issue. Winograd is to be praised for identifying the acceptance problem, but not to be praised for his narrow view of inductive methods. Incidentally, in the revised, *AI Journal* version of this remark [Win80b], Winograd attacks "mathematical logic" instead.

Going back to Reiter's closed world assumption paper, one discovers an amazing historical fact. Reiter was originally interested in a genuinely semantical problem, and the interest in semantics must have persisted as he turned to inferential problems. Reiter disclaimed an interest in inference; he was concerned with representation:

> . . . the number of negative facts about a given domain will, in general, far exceed the number of positive ones so that the requirement that all facts, both positive and negative, be explicitly represented may well be unfeasible. [Rei78a]

He enjoined us, "merely *explicitly* represent *positive* facts"; negative facts were to be implicitly represented. "*The implicit representation of negative facts presumes total knowledge about the domain being represented* . . .; fortunately, in most applications, such an assumption is warranted." So if it was the case that neither p nor ~p was explicitly represented, then it was the case that ~p was known and was implicitly represented. If p later became explicitly represented, this was outright revision: the revision from ~p to p; not non-monotonic inference. It couldn't be non-monotonic inference (or monotonic inference, or any kind of inference) since there was no

inference. The closed world assumption was an inference, perhaps inductive, but that was an inference made by the data base designer, not by the inference engine (what inference engine?).

In the original formalism, belief fixation and revision pretended to be nothing more than unexplained change in the set of represented beliefs. Reiter had no reason to look at the philosophers' work on inferential mechanisms for beliefs.

In default logic [Rei78b,80], some belief revision was inferential: change that was explainable in terms of inferences that were defeated. The interesting problems of default reasoning weren't semantic; everyone agreed what beliefs were represented (implicitly or explicitly) by a set of sentences in the language. Unfortunately, with the closed world "assumption" already in the air, it was more important to relate default reasoning to A.I.'s closed world assumption than to the defeasible reasoning styles in another discipline.

Doyle and McDermott's contribution to the split was a matter of dubious ambitions. In the TMS paper, Doyle clearly recognizes that truth and acceptability should not be confused:

> Briefly put, the problems with the conventional view of reasoning stem from the monotonicity of the sequence of states of the reasoner's beliefs: his beliefs are true, and truths never change . . .. To study rational thought, we should study justified belief or reasoned argument, and ignore questions of truth. [Doy79]

He talks about the "philosophical literature [that] includes many treatments of belief revision and related problems." One wonders whether David Israel would have complained the way he did if he had focused on the TMS paper instead (Doyle has actually said something like this [Doy85]). It was clear to Doyle that TMS and non-monotonic logic could be embedded in the philosophical tradition. McDermott and Doyle rehearsed the idea that non-monotonic logic fit into a framework that included belief revision, logic of counterfactual conditionals, and "world-model reorganization," citing Quine and Ullian, Rescher, and Scriven. If not for the choice of presentation, Doyle and McDermott would clearly be talking about real rules.

But in order to provide "theoretical foundations," Doyle and McDermott represent belief by assertion, thus taking all of deductive systematization's rules as belief fixation's rules. They take the Lehrer-Levi-Hintikka-Hilpinen position for granted. And they treat consistency as a proposition-forming modality. They wanted to add to deductive logic, not subtract from it. The obvious alternative, which had always been used in the philosophical literature, was to discuss consistency and inferential relations in a meta-language.

Surely they knew they could have gone the meta-linguistic route. Careful reading again shows that their choice was an informed one. The TMS paper mentions the meta-theoretic approach of Weyhrauch right next to the modal approaches of Hintikka and Moore. The joint paper notes the meta-linguistic approach of Kramosil. They are disappointed by Kramosil's "pessim[ism] with regard to the possibility of formalizing [non-monotonic] rules" and they export the

disappointment to the whole meta-linguistic approach he uses: "Non-monotonic inference rules need not appear in the explicit forms discussed by Kramosil." Furthermore, McDermott and Doyle seem bent on analyzing default locutions in a structure-mimicking form. So

> if something is an animal with a beak then unless proven otherwise,
> it is a bird,

must be composed of constituents like

> if [animal x & has-beak x] unless [|- ~bird x]
> then [bird x],

entirely in the object language. One alternative is to say

> animal x & has-beak x is a prima facie reason for bird x

and simply leave it at that. Apparently it was the McCarthy and Hayes paper that influenced them here; McDermott and Doyle refer to it for an explanation of the modality M. McCarthy and Hayes proposed that problems related to frames could be handled by introducing modal operators, and they added a warning, "We hereby warn the reader, if it is not already clear to him, that these ideas are very tentative and may prove useless." The MIT AI lab duo may simply have taken this disclaimer to be an irresistible challenge.

Perhaps because of the success and prominence of the McDermott and Doyle work, A.I. never returned to the meta-linguistic approach, and some philosophers have been quietly pointing and laughing at our A.I. ilk ever since.

What about the pointing and laughing? Apparently what resulted was pretty good, despite its presentation or history. The prominent philosophers Glymour and Thomason even applaud the uninhibited steps:

> Artificial Intelligence has done us the service not only of reminding us of the importance of non[-]monotonic reasoning, but of demanding a qualitative, logical account of it, and of suggesting how such an account might be formulated.
>
> There is room for creativity and judgement here, and we can naturally expect different solutions to evolve. [GlT84]

A.I.-style non-monotonicity really was a coup in knowledge philosophizing and engineering. Someone notice that even with great simplifications in formalism, something like a conditional logic could reproduce most of the characteristics of inductive reasoning. Like acceptance, qualitative defeasible reasoning also produces infallible, corrigible statements for reasons.

Take Doyle's epistemological programme, for example. His agent would declare defeasible policies in advance, and they would be time-invariant. Doyle focused not on inferring the policies, but on inferring with them. This is all that is needed when the agent is

ideally cooperative and declares all the dependencies among beliefs of interest, rationally or otherwise. The agent would shift subjectively from state to state, report partial descriptions of the new state, and Doyle would help fill out the description by following the invariant policies.

Contrast this with inductive logic: the evidential base of the agent changes from state to state (it grows) and is reported; then a *general* principle completes the description of what must be the beliefs rationally held in light of this evidence. This general principle is an acceptance rule in a traditional form, e.g., "accept p iff the probability of p relative to blah blah is blah high blah and p coheres/explains blah blah blah." This acceptance rule is presumably invariant like TMS's invariant policies. But it doesn't look like any of TMS's policies. TMS policies are particular; they contain predicate and term names of the object language, and there's a long list of them. Theoretically, the invariant CP and SL statements that encode defeasible policies in TMS could be derived from the general principle for acceptance. If so, the derived CP and SL statements would be no arbitrary set; they would have those logical properties that are characteristic of the acceptance rule (such as the consistency of k-membered subsets).

Still, the distinction is subtle.

Inductive logic was slightly better for use in revision because with it and the current evidential corpus, all warranted defeasible policies were determined. K may not be specified completely, but the policies warranted with respect to what is specified are completely determined (though not enumerable because of undecidability). Doyle needs to supply his policies one by one, and he pays the price when the defeasible policy needed isn't supplied. He won't know whether a conclusion couldn't be drawn (or decided) or whether a conclusion could be drawn but its rule simply wasn't stated. There's always a nagging suspicion that the missing rule could have been supplied. How big a problem this is depends on what you think about the relation between computation, information, error, decision, and unsound inference, which I elaborate below. Let's just say that not everyone considers the closed world assumption for CP and SL statements (or default rules, or conditionals, etc.) compelling in practice. Inductive logic doesn't need an explicit list of policies or prescriptions. Confirmation theory always says which among the alternative rules is better confirmed and by how much; acceptance rules tirelessly say whether the "how much" is "enough." Since they are inherently quantitative, they can also be made sensitive to various attitudes toward error and demands for information.

This is what Doyle and the rest have given up. But there may be nothing else of importance that has been sacrificed; for this price, A.I. buys the ability to leave quantitative confirmation theory and its messy issues of statistics, decision, and measurement. Clearly a possible line of future research is the pursuit of default systems with more structure or more inference rules, so that the closed world assumption for these defaults can be widely accepted, and the boldness of inference can be sensitive to the demands for information.

To sum at this point, we should agree with Israel that the

presentation of reasoning systems in the past has been confused, and agree that there is much to be learned from a serious look at the post-Carnapian philosophical literature. In the same breath, we should also agree that pursuing qualitative, defeasible, real rules of inference is a promising line of research that is independently motivated.

Folk Myths.

To Glymour and Thomason, inductive inference is a subspecies of non-monotonic reasoning:

> Philosophers and logicians have of course payed attention to [non-monotonic] reasoning, which includes not only ceter[i]s paribus reasoning, but all forms of inductive and statistical inference. But it has been fashionable to treat such matters as fundamentally quantitative, by subsuming them under probability theory. [GlT84]

They don't say what is the relation between inductive inference and *rational* non-monotonic reasoning. I have left open the question of whether all rational non-monotonic reasoning is subsumed by the epistemologists' inductive reasoning, broadly construed. If there is a dispute here, it probably depends on whether you classify the rules for theory formation as inductive, and on how you view rationality.

I want to focus on two related but different claims. Both challenge the autonomy of non-monotonic reasoning and both are clearly false. The first is that non-monotonic reasoning can be reduced to probabilistic reasoning. If rational, it may or may not reduce to *induction; but it certainly doesn't reduce to probability.* The second is that induction or probability "solves" the multiple extension problem. It solves decision problems that may relate to the multiple extension problem, but inferentially, it does nothing that default logic, say, couldn't do. I think the claims express sentiments that are true when worded precisely, but that as they stand, they are dangerously misleading folk myths.

Induction is of course what many have had in mind when they've championed probability theory as the cure for non-monotonic reasoning's woes. The confusion of induction and probability is due to the popularity of Bayesian decision theory as a means of getting along without a rule of acceptance. Inductive logics are characterized by their confirmation theory and their acceptance rules. Supposing that there are no acceptance rules and that confirmation is just probability makes probability look like an inductive logic. But it is induction, if anything, that is the candidate for subsuming rational non-monotonic inference.

Some think acceptance requires more than probability. Levi uses cognitive utilities (so do Hempel [Hem62] and others). Real-valued measures have been used for confirmation, which do not satisfy finite additivity (see [Kyb64]).

Even if confirmation is just a matter of high probability, probability by itself does not lead to non-monotonic inference. Unless statements are accepted, they aren't inferred; there is just a

calculation of probabilities. Having probability -- whatever the value -- is not the same as having been inferred. Non-monotonicity refers to the set of statements that may be inferred, as evidence grows; it does not refer to evolution of probability quantities, except when they lead to inference. The only kind of inference that's non-monotonic in the Bayesian scheme is the inference that a particular decision is optimal. Probability couldn't generate this non-monotonic behavior without the decision theory; it's the decision theory that leads to the acceptance of the statement of optimality. Probability doesn't subsume non-monotonic inference unless it combines with an honest-to-goodness rule of acceptance. If there is such a rule, one that introduces corrigibility and infallibility, then it's an inductive logic.

An interest in probability in A.I. is undoubtedly a good thing; probability is potentially useful to A.I. for reasons other than subsumption, including its importance in confirmation and in decision-making under risk. What probabilists should say is that whenever practitioners have applied non-monotonic reasoning, they could have used probabilistic reasoning, with some decision theory.

McCarthy speaks of circumscription as a "streamlined expression of probabilistic information when numerical probabilities, especially conditional probabilities, are unobtainable." [McC84] We must remember that even if the conditional probabilities were obtainable, we still wouldn't be able to duplicate the reasoning without a theory of acceptance. It's more aptly a streamlined expression of inductive information. Elaine Rich explains "If we ask what default reasoning really is, we see that it is a form of likelihood reasoning." [Ric83] We must remember that she has in mind a non-trivial connection between default reasoning and likelihood. Cheeseman [Che85] says that the default "All birds fly unless proved otherwise" should really be "Most birds fly," where the latter is "used as a piece of evidence in evaluating the probability of the proposition 'this bird flies' . . .." Default rules in non-monotonic inference engines are like conditional probabilities in probabilistic inference engines. But that doesn't mean defaults can simply be reduced to conditional probabilities. People may "force intrinsically probabilistic situations into a logical straight-jacket," but it is equally wrong to force intrinsically inductive situations into the straight-jacket.

As for multiple extensions, appeal to an inductive logic in a weak sense preempts discussion of what can or should be done when there are multiple extensions. In principle, the logic says exactly what is and isn't acceptable on current evidence. Note that this property of the logic depends on the acceptance rule; if the high probability rule had been ambiguous between Acc(h) and not(Acc(h)) when

there are Qi, Qj in Q s.t. $Qi(h) < 1 - e < Qj(h)$,

then this wouldn't be the case. Undecidability of the acceptance rule also tempers the claim that inductive logic fares better than non-monotonic A.I. systems. Suppose ambiguity and decidability of the acceptance rule aren't at issue.

Still, whenever the problem can be solved by induction or by using probability, it can also be solved within the framework of

default or non-monotonic logic. Consider the simple case:

A : MC |- C
B : M~C |- ~C
A & B & K.

If induction says that C is acceptable on A & B & K, then there is nothing peculiar to the *formalism* that prohibits the rule

A & B & K : MC |- C,

which presumably defeats the rules with less specific antecedents. Epistemologically and practically speaking, we don't expect that the rule will always be present; it would require an infinitude of default rules to reflect the behavior of the inductive logic. Nevertheless, multiple extensions are a problem for non-monotonic reasoning systems only because the required rule might not be declared, and the closed world assumption not assumed. With the right rule present, there's no problem.

If induction "solves" the multiple extension problem, so does the presence of adequately specific defaults. The claim that induction or probability "solves" multiple extension problems is either a trite claim, or a myth.

What is less trite is the idea that unsound inference is associated with decision. Induction and probability are better mates for decision theory than any of A.I.'s existing non-monotonic systems. Apparently the multiple extension problem, the demand to know whether C or ~C (or neither, or both) looks like a demand for information, or a decision. If these ideas are right, they explain why inductive logic and probability theory appear to handle multiple extension situations better.

Decision rules based on logical or personal probability make explicit use of degrees of belief. Acceptance rules used to construct particular bodies of knowledge use parameters appropriate to the pragmatic considerations of informational demand and attitude toward error. Both kinds of rules are sensitive to the stakes. Each in its own way allows variation of the boldness of its inferences. Each can respond differentially to various practical demands. "C is inferrable on A & B & K if it matters *this* much to you." In contrast, the boldness of inference in an A.I. system's fixed set of defeasible rules is implicit and immutable. These rules cannot be made to respond to the particular decision at hand; C is either inferrable on A & B & K or it isn't, now and forever. Here's an analogy: the Bayesian view of confirmation in science presupposes there are fixed utilities for objective, scientific inquiry. Utilities can't be changed relative to objectives. Confirmation for one is confirmation for all. The result: a whole dimension of critical control is lost.

Whether it's a feature or a flaw is not the main issue; clearly people who need an inference mechanism with variable boldness are free to forego existing non-monotonic systems. In short, they're free to use decision theory on decision problems. It's also straightforward to augment A.I.'s non-monotonic systems so that they are sensitive to varying

practical demands, so that inference can be made cautious to degrees [LFK85].

The main point is that there's something to the folk myth, even though it takes a mouthful to say it right. And it's now possible to say why the closed world assumption is so bad for defeasible rules. When creating knowledge bases, there is the problem of writing down particular defeasible rules, each with its own boldness or propensity to err. Unless the boldness of a system's defeasible inference can be set a priori, univocally, and appropriately for all future inquiry and action, it's not clear which rules are warranted and which are not. Stipulation is fine. But when we legitimately demand to know whether C or ~C (or neither, or both), the prior stipulation to exclude A & B & K : MC |- C on insufficient warrant may be called into question.

Perpetrators of the myth should say something like this: real rules of inference are sensitive to tradeoffs between desiderata of information and error.

Are We Interested in "The Epistemological Problem"?

No doubt, we are all interested in the epistemological aspects of non-monotonic reasoning problems, rather than the semantical or proof-theoretical aspects. So much for who is interested in Israel's epistemological problem.

What about the acceptance problem that the philosophers are so willing to give on loan? It will be interesting to those who will represent Q, including the evidential reasoning people. Even avid Bayesians like Cheeseman recognize that pure probabilism commits its adherent to multiplicatively large measure spaces or algebras. Eliciting, manipulating, and maintaining probability measures over such spaces is impracticable computation.

Abduction, presupposition, and convention place inquiry in deductive channels that are fruitful and efficient. Induction does too, and unlike the others, it is eminently accountable to rationality.

Inference that leads from probability to infallibility is desirable because it allows subsequent inference that is informationally effective and computationally efficient. It permits conditioning on statements that are not strictly observational. It perhaps even permits familiar heuristic deductive inference. It permits the acceptance of (x).Ax -> Cx. Taking the generalization to be infallible allows it to be used as if it were theoretical or analytic knowledge, which is good; it is what Brachman has fussed about [Bra85] (though Brachman seems to want generalizations that are incorrigible, too, and doesn't really care whether they're inferred). The related statistical generalization, P(Cx | Ax) > 1 - e, is less informative. With the infallible generalization, ~Ca |- ~Aa, and Aa & ~Ca can be omitted from serious possibility in inquiry.

Whenever the generalization is so confirmed that it is acceptable, it must be that treating the counter-examples to the generalization, the long shots, as serious possibilities is just a nuisance. In these cases, treating them and not treating them both lead to the same decisions. A view of this kind, which worries about the (computational) limitations of practical deliberation, is developed

by Harsanyi [Har85]. From this view, the acceptance problem is indeed interesting to practical evidential reasoning.

The more various the inferential rules governing infallibility, the more the criteria for the rationality of the corpus. When relevant evidence for an infallible statement can be completely specified, acceptance rules serve as a coherence constraint. Statements had better be acceptable on represented evidence, lest there be legitimate demand for external justification. Deductive consistency is often a coherence constraint demanded of knowledge bases. Acceptability from evidence may serve as a constraint too, at least for special classes of statements in knowledge bases, such as empirical generalizations. So designers of coherent knowledge bases and careful eliciters of expert knowledge should care about acceptance.

Acceptance will be interesting even to those who do not presently use Q. Consider Mitchell's learning work [Mit83]. Integration problems take various forms, depending on the functions used in the integrand. Forms can be transformed by operations, such as separation by parts. Mitchell's program infers from past experience that particular operations are best on particular forms. Whenever each of two operations independently appears to be best, the program attempts to divide the form into subforms, so that operations are unambiguously best on their subforms.

Unfortunately, the language may not provide any way to divide a form, in such a way that the best operation becomes unambiguous; the best operation may be ambiguous on a primitive form. Presently Mitchell can choose one operation based on recent experience, or leave the ambiguity and underdetermine the solution strategy for the form in question. There is an evidential alternative: take previously encountered problems as empirical data, from which future problem character can be inferred. So "trigonometric integrand" may be a primitive, indivisible form. OP1 and OP2 have each been observed to be best on this form, on various respective occasions. If the evidence in favor of OP1 is sufficient to accept that OP1 is unequivocally best for that form, the action taken on trigonometric integrands can be simplified. If the action taken when there is ambiguity is already simple, such as rolling a die, nothing has been achieved. But if ambiguity leads to special, information-seeking behavior, such as experimentation, there will be a payoff. OP1 may be so overwhelmingly acceptable that further analysis is a waste of time.

Symbolic learning in A.I. has generally been conceived as induction of a statement, criterion, or procedure based on inconclusive evidence. Research on learning has adopted induction in name; one sees the work referred to as induction; but implementers have not seriously tapped the philosophers' work. There's an explanation: symbolic learning in A.I. to date has been a search for the simplest account, in some language, that fits given data perfectly. The paradigm example is Winston's program [Win75], which looks for the shortest description of an arch that correctly distinguishes all of the externally supplied examples. He presupposes that there is such a description.

As soon as learning problems become so complex that no description in the supplied language corroborates with the data

perfectly, as soon as there is no choice but to quantify error, acceptance will become interesting. Also, when autonomy is increased, when there is a decision to obtain data, rather than just receive it whenever it pleases the teacher, confirmation and acceptability will become an issue.

Acceptance generally addresses problems of independent knowledge acquisition. Acceptance just says what knowledge can be acquired on the basis of what. In their work on conditional logic, Glymour and Thomason note:

> For a system that approximates our logical theory to work successfully, it must have many default rules, expressed by conditionals. And it is clear that these conditionals have to be *generated* somehow, rather than simply listed. [GlT84]

They announce the generation of conditionals as a forthcoming attraction. They are caught in the same position as Doyle and the rest of the qualitative non-monotonicity researchers. General acceptance rules can generate defeasible policies, which otherwise have to be supplied externally. But their system doesn't use a general acceptance rule. It relies on external suppliers.

Why shouldn't all knowledge simply be supplied? Something has to be supplied externally. Inference to the infallibility of contingent statements can't be done in a vacuum. Epistemological problems in A.I. always come with a class of statements that is "directly observational." It is convenient to consider a certain class of potential statements infallible and incorrigible, as evidence for the acceptability of less privileged statements. These infallible, incorrigible statements are the observation reports: e.g., "pixel on at line y, column x," "force at tip measures f mg/s↑2," or "apparently an object at time t and position p."

The conceptual level of observation reports is several abstractions below the level presumed by much of the A.I. work on inference. Something has to bridge these levels. That's the work done by acceptance. The alternative is to represent probability measures over all levels simultaneously.

The philosophers have their own reasons for acceptance; they have primarily to do with the formalization of inference. At some point, someone has to write something down. C.I. Lewis, for instance, thinks "If anything is to be probable, then something must be certain." [Lew46]. But the value to A.I. is definitely in the bridging.

With good bridging, knowledge bases can be founded on empirical data in an effective, systematic, and rational way. Undoubtedly, comprehensive reasoning systems (if any are ever built) will be based on local induction.

Epistemological Eyes.

Must we avert our epistemological eyes? A.I. has yet to look at the epistemological foundations of knowledge bases.

How do statements come to be regarded as infallible? Which statements are to be corrected in light of additional given evidence?

What is the nature of the commitment to infallible statements; i.e., what good is commitment if it is held corrigible, and what is done with infallible statements that can't be done with statements qualified by their degrees of confirmation? There should be answers to these questions and they should be principled. A.I. workers seem to be operating without awareness of these questions, and without any answers.

I have argued that a major failure for A.I. has been a failure to appreciate the work on inductive logic and epistemology. Of course the traditional epistemological topics (skepticism, justification of induction, and semantic analysis of knowledge) have proven to date mere impractical philosophical primping. But that's not all there is. Induction is construed broadly these days: not merely the inference to the general from particulars. It is considered the unique problem in philosophy the answer to which bears on the problem of legislating rational belief. Contemporary epistemologists following Carnap's lead have been constructing formal rules for the acceptance of statements and expounding the foundations of probability and confirmation. This is the work that should be interesting to A.I.. Oddly, none of this appears in the classic old "Epistemological Problems of A.I." [McC77], or in the suspect new "Intelligent Systems and Applied Epistemology," [vDa85] or in any other A.I. paper with "epistemology" in the title in between. It was disappointingly absent from Dennett's correspondence reports, too [Den82,84].

A.I. researchers familiar with probability and with logic from mathematics and engineering backgrounds (or worse) should understand that there is a discipline mediating the two, and that it is this discipline which is best suited to interact with A.I. in the study of reasoning. Winograd's remark is irrefutable evidence of the need for understanding.

A.I. work on inference has chosen to ignore the acceptance problem, while work on evidential reasoning has either considered acceptance inessential, or displayed ignorance of its pitfalls. We have proceeded in the past without visible epistemological foundations by way of declaration. The contents of knowledge bases have been merely declared, and so have the policies for revision. BLOCK7 has borne the ON-relation to TABLE by declaration. The ability of TWEETY to FLY has yielded in the presence of the statement that TWEETY is a CARTOON-CHARACTER, by declaration. What has been said is that declaration of knowledge is unprincipled, and that any principled alternative is going to be inductive, which could require that deduction play second seat.

If A.I. were to draw more heavily from the work in inductive logic, especially the work on acceptance, it would not have to proceed solely by declaration. A more principled way of proceeding would allow automated construction and maintenance of large knowledge bases. Israel and I have also claimed that appreciation of inductive logic and related work on defeasible reasoning solves conceptual puzzles that have been mentioned in the A.I. literature. Because this appreciation clears conceptual confusion, it will result in research effort directed in the appropriate places, with the right formalisms and the right context. It will also suggest entirely new ways of implementing non-monotonic inference systems. Philosophers in turn will benefit

from dialogue, motivation and feedback: from the observations of the system's synthetic performance; and from A.I.'s help on some of the heuristics needed in practice.

Israel said to go to the philosophers. I have been to the land of the philosophers of science and this is what I have seen.

Acknowledgements.

This paper has been prepared under a grant from the U.S. Army Signals Warfare Laboratory.

This interdisciplinary work is the result of some vision on the part of Henry Kyburg and Pat Hayes. Kyburg thought it would be a simple matter to relate the two literatures, but never realized how guarded the claims had to be, especially since one of the main figures responsible for the divergence was Hayes!

[Asc84] Ascher, N. "Linguistic Understanding and Non-Monotonic Reasoning," AAAI Workshop on Non-Monotonicity 1984.

[Bra85] Brachman, R. "I Lied About the Trees," AI Magazine 6, 1985.

[Car50] Carnap, R. Logical Foundations of Probability, Chicago, 1950.

[Car62] Carnap, R. "The Aim of Inductive Logic," in Logic, Methodology and Philosophy of Science, E. Nagel, et. al., ed. Stanford Univ. Press, 1962.

[Che85] Cheeseman, P. "In Defense of Probability," IJCAI 1985.

[Den82] Dennett, D. "Recent Work in Philosophy of Interest to A.I.," AI 19, 1982.

[Den84] Dennett, D. "Recent Work in Philosophy II," AI 22, 1984.

[Doy79] Doyle, J. "A Truth Maintenance System," AI 12, 1979.

[Doy83] Doyle, J. "Methodological Simplicity in Expert System Construction," AI Magazine 4, 1983.

[Doy85] Doyle, J. personal communication.

[EtR83] Etherington D. and Reiter, R. "On Inheritance Hierarchies With Exceptions," AAAI 1983.

[FaH85] Fagin, R. and Halpern, J. "Belief, Awareness, and Limited Reasoning: Preliminary Report," IJCAI 1985.

[GlT84] Glymour, C. and Thomason, R. "Default Reasoning and the Logic of Theory Perturbation," AAAI Workshop on Non-Monotonicity 1984.

[GMc84] Good, I. and McMichael, A. "A Pragmatic Modification of

Explicativity for the Acceptance of Hypotheses," Phil. Sci. 51, 1984.

[Har85] Harsanyi, J. "Acceptance of Empirical Statements: A Bayesian Theory without Cognitive Utilities," Theory and Decision 18, 1985.

[Hay85] Hayes, P. personal communication.

[Hem62] Hempel, C. "Deductive-Nomological versus Statistical Explanation," Minnesota Studies in the Philosophy of Science 3, Minnesota, 1962.

[HiH66] Hintikka, J. and Hilpinen, R. "Knowledge, Acceptance, and Inductive Logic," in Hintikka, J. and Suppes, P., eds., Aspects of Inductive Logic, North-Holland, 1966.

[Hil68] Hilpinen, R. "Rules of Acceptance and Inductive Logic," Acta Philosophica Fennica 22, 1968.

[Isr80] Israel, D. "What's Wrong with Non-Monotonic Logic?", AAAI 1980.

[Jef70] Jeffrey, R. "Dracula Meets Wolfman: Acceptance versus Partial Belief," in Swain, M., ed. Induction, Acceptance, and Rational Belief, Reidel, 1970.

[Kle85] Klein, P. "The Virtues of Inconsistency," Monist 68, 1985.

[Kyb64] Kyburg, H. "Recent Work in Inductive Logic," American Philosophical Quarterly 1, 1964.

[Kyb70] Kyburg, H. "Conjunctivitis," in Swain, M., ed. Induction, Acceptance, and Rational Belief, Reidel, 1970.

[Kyb74] Kyburg, H. The Logical Foundations of Statistical Inference, Reidel, 1974.

[Kyb83] Kyburg, H. "The Reference Class," Phil. Sci. 50, 1983.

[Leh70] Lehrer, K. "Justification, Explanation, and Induction," in Swain, M., ed. Induction, Acceptance, and Rational Belief, Reidel, 1970.

[Leh74] Lehrer, K. Knowledge, Oxford, 1974.

[Lev80a] Levi, I. The Enterprise of Knowledge, MIT Press, 1980.

[Lev80b] Levi, I. "Potential Surprise," in Cohen, L. and Hesse, M., eds., Applications of Inductive Logic, Oxford, 1980.

[Lew46] Lewis, C. An Analysis of Knowledge and Valuation, Open Court Publishing Company, 1946.

[LFK85] Loui, R., Feldman, J., and Kyburg, H. "Interval-Based Decisions for Reasoning Systems," Proceedings of the AAAI Workshop on Uncertainty and Probability in Artificial Intelligence, 1985.

[McC77] McCarthy, J. "Epistemological Problems of Aritificial Intelligence," IJCAI 1977.

[McC84] McCarthy, J. "Applications of Circumscription to Formalizing Common Sense Knowledge," AAAI Workshop on Non-Monotonicity 1984.

[McD80] McDermott, J. and Doyle, J. "Non-monotonic Logic I," AI 13, 1980.

[McD85] McDermott, J. "Easy and Hard Problems in Artificial Intelligence," address to the Artificial Intelligence Society of New England, Brandeis, November 1985.

[McG85] McGee, V. "A Counterexample to Modus Ponens," J. Phil. 35, 1985.

[McH69] McCarthy, J. and Hayes, P. "Some Philosophical Problems from the Standpoint of Artificial Intelligence," Machine Intelligence 4, 1969. Reprinted in Webber, B. and Nilsson, N., eds. Readings in Artificial Intelligence, Tioga, 1981.

[Mit83] Mitchell, T. "Learning and Problem Solving," IJCAI 1983.

[Noz81] Nozick, R. Philosophical Explanations, Harvard, 1981.

[Nut84] Nute, D. "Logical Relations," Philosophical Studies 46, 1984

[Nut85] Nute, D. "A Non-monotonic Logic Based on Conditional Logic", working paper, Center for Advanced Computational Methods, Atlanta, 1985.

[Pol74] Pollock, J. Knowledge and Justification, Princeton, 1974.

[Pol83] Pollock, J. "A Theory of Direct Inference," Theory and Decision 16, 1983.

[Qui83] Quinlan, J. "Consistency and Plausible Reasoning," IJCAI 1983.

[Rei78a] Reiter, R. "On Closed World Data Bases," in Logic and Data Bases, Gallaire, H. and Minker, J. eds., Plenum, 1978. Reprinted in Webber, B. and Nilsson, N., eds. Readings in Artificial Intelligence, Tioga, 1981.

[Rei78b] Reiter, R. "On Reasoning by Default," Proceedings TINLAP, Urbana-Champaign, reprinted in Brachman, R. and Levesque, H., eds. Readings in Knowledge Representation, Morgan-Kaufman, 1985.

[Rei80] Reiter, R. "A Logic for Default Reasoning," AI 13, 1980.

[ReM70] Rescher, N. and Manor, R. "On Inference From Inconsistent Premisses," Theory and Decision 1, 1970.

[Ric83] Rich, E. "Default Reasoning as Likelihood Reasoning," AAAI 1983.

[Sta85] Stalnaker, R. Inquiry, MIT Press, 1985.

[SzP78] Szolovits, P. and Pauker, S. "Categorical and Probabilistic Reasoning in Medical Diagnosis," AI 11, 1978.

[vDa85] van Damme, F. "Intelligent Systems and Applied Epistemology," Communication and Cognition - AI 2, 1985.

[Win75] Winston, P. "Learning Structural Descriptions from Examples," The Psychology of Computer Vision, McGraw-Hill, 1975; reprinted in Brachman, R. and Levesque, H., eds. Readings in Knowledge Representation, Morgan-Kaufman, 1985.

[Win80a] Winograd, T. "Extended Inference Modes in Reasoning by Computer Systems," in Cohen, L. and Hesse, M., eds., Applications of Inductive Logic, Oxford, 1980.

[Win80b] Winograd, T. "Extended Inference Modes in Reasoning by Computer Systems," AI 13, 1980.