COGNITIVE SCIENCE **11**, 283–297 (1987)

Response to Hanks and McDermott: Temporal Evolution of Beliefs and Beliefs about Temporal Evolution

RONALD P. LOUI
University of Rochester

This paper critically evaluates the celebrated paper of Hanks and McDermott on temporal projection, non-monotonic reasoning, and the frame problem. First I argue against their intuitions, and a fortiori, against their proposed solution. Next, I suggest how the solution they desire could be obtained, were they willing to represent the problem a bit differently.

I. THE HANKS-McDERMOTT PROBLEM

Steve Hanks and Drew McDermott (1986) describe a temporal projection problem that they believe (1) exhibits an important kind of reasoning for practical Artificial Intelligence systems and (2) cannot be handled by the existing non-monotonic inference systems. I don't share either of these beliefs. In this note, I will point out why I am neither bothered by their temporal projection problem nor convinced by their analysis.

The problem involves reasoning about a gun, known to be loaded at a time, and fired at a person at a later time. We want to know if the person ceases to live. We are willing to assume that if the gun remained loaded, then the firing was effectively fatal. But the kind of reasoning that permits us to conclude that the gun remained loaded until it was fired would also appear to allow reasoning to the conclusion that the person remained alive, even after the firing. One choice is to reason that the property of being alive persists, and hence, the property of being loaded does not. The other choice is to reason that the property of being loaded persists and the property of being alive does not.

Symbolically, in an impoverished temporal representation (but one that is adequate for our purposes), we have (see Figure 1):

As usual, discussion with Henry Kautz and Rich Pelavin was fruitful. Rich Thomason provided useful criticism.

Correspondence and requests for reprints should be sent to Ronald P. Loui, Departments of Computer Science and Philosophy, University of Rochester, Rochester, NY 14627.

loaded@0
alive@0
fired@1
(1) if fired@1 and loaded@1 then not-alive@2
(d1) if alive@0 then defeasibly alive@1
(d2) if alive@1 then defeasibly alive@2
(d3) if loaded@0 then defeasibly loaded@1

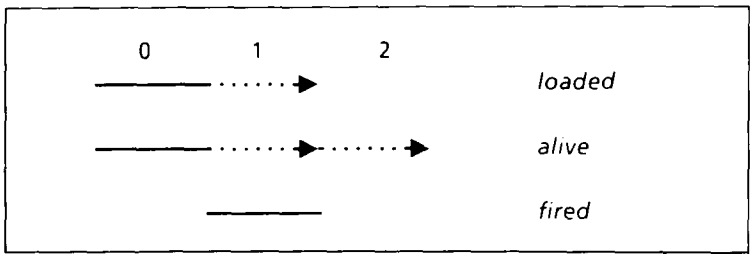

Figure 1.

Hanks and McDermott think it is intuitive that an adequate defeasible reasoning system will allow not-alive@2 to be concluded, i.e., that the property of being loaded persisted and therefore the property of being alive did not. They correctly note that there is a temporal asymmetry between (d2) and (d3). Namely, (d3) refers to earlier times. Hanks and McDermott next propose that defeasible conclusions be ordered according to a "temporally forward" priority. Earlier defeasible conclusions defeat later defeasible conclusions, if the conclusions are contraries. They mention the work of a close colleague, Yoav Shoham (1986), who justifies this approach on the basis of reflections on causality.

So loaded@1 and alive@1 can be concluded, but alive@2 cannot be, because we have already committed ourselves to loaded@1. In the choice between loaded@1 and alive@2, the former is preferred.

II. IS THE INTUITION COMPELLING?

The reasoning situation is not a problem for existing non-monotonic or defeasible reasoning systems if one rejects the Hanks-McDermott intuition that not-alive@2 is the mandatory, albeit defeasible, conclusion.

Why should we believe not-alive@2? Surely we can believe that loaded@1, therefore, not-alive@2. Or else we can believe alive@1, therefore alive@2, therefore not-loaded@1. It's true it's odd to say that being alive at a time caused the gun to be unloaded at an earlier time. But we are not committed to saying that. Causal laws may be involved, but the reasoning need not be from causes to effects. Moreover, *the event* of my inferring that the subject is alive will belong to the causal chain that leads to *the event* of my inferring

that the gun was unloaded. These *events* of inferring are correctly ordered temporally. It's not necessary to violate temporally forward conventions of causality in order to hold alive@2.

Hanks and McDermott most likely have in mind the forward-chaining part of a planner. It is supposed to reason that the system can achieve its goal of not-alive@2 by performing an act, the effects of which guarantee fired@1. Isn't it just plain desirable to reason that firing at time 1 will achieve the goal? It is, if at least one of the following is true:

(a) actions can be performed that (acceptably) guarantee that the gun will not be unloaded between times 0 and 1; or
(b) the possibility of unloading between times 0 and 1 is not considered a serious possibility by the planner.

If we assume (a), then Hanks and McDermott have no problem. Just cite those actions (that guarantee loaded@1 if loaded@0) as part of the plan. (b) is more interesting. Rich Pelavin's dissertation (Pelavin, 1986) discusses the problem of "airtight" planning in non-deterministic worlds. His relevant contribution here is the idea that we must antecedently define the possibilities we are willing to consider, of those that could subvert our plans. If Hanks and McDermott want to consider explicitly the possibility of unloading, then they cannot expect the planner to reason that firing at time 1 will be sufficient. On the other hand, suppose they do not want to consider explicitly the possibility of unloading. Then the is-loaded persistence rule should not be defeasible. In either case, either the problem is misrepresented, or there is no problem.

As the problem is stated, there's nothing wrong with concluding that there could have been an unloading, hence alive@2.

If we believe that not-alive@2 is mandatory for a defeasible inference system, or even desirable, given the knowledge that is explicitly represented, then we are apparently committed to some very undesirable inferences, in related situations. Consider a business school student known to be registered for full-time studies at some time.

Example II.1. (see Figure 2)

MBA-student@1
(d1) if MBA-student@1 then defeasibly MBA-student@2
(d2) if MBA-student@1 then defeasibly MBA-student@0
(1) if MBA-student@0 then not-MBA-student@2.

If (d1) is a "forward persistence axiom," then (d2) is a "backward persistence axiom." It seems in this example that backward persistence is as desirable, as probable, and as warranted as forward persistence. And it doesn't matter whether we are engaged in historical reasoning or in presentiment. A defeasible rule such as (d2) could be important in predictions about

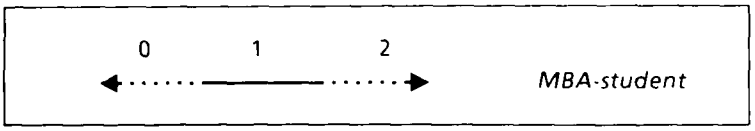

Figure 2.

the future as well as the past. It may be, for instance, that what is at stake is future salary, which depends on when our MBA student actually graduated.

Of course, rules like (d1) and (d2) must be justified, and there may be reason to prefer one over the other. Unfortunately, there isn't much discussion to be found on the justification of defeasible rules (see Loui, 1986b). Still, whatever justifies (d1) is likely to justify (d2) in this example, whether the justification is semantic, or has a habit of inference, or as a high probability association.

The problem with Hanks and McDermott's and Shoham's forward-marching solutions here is that they conclude MBA-student@0, and not-MBA-student@2. Given that we know our young corporate aspirant is in school, we are obliged to conclude that he or she is on the verge of graduation![1]

Examples can be found at will that share exactly the same syntactic structure of the Hanks-McDermott problem, but do not seem to require the analogous conclusion. It does not seem to matter whether the laws involved are nomological, or reflect causal relations, or are just laws of association. In the two examples that follow, the strategy of drawing defeasible conclusions in a way that prefers earlier conclusions is repeatedly shown to be lacking. In fact, in the latter of the two examples that follow, the conclusion that is analogous to concluding alive@2 is the intuitively desirable one, contrary to Hanks and McDermott's pattern of reasoning.

Consider the extinction of the South American marsupial carnivore, Thylacosmilus. We don't know whether Thylacosmilus had exclusively South American extent during the entire Teritiary period; we know at the early Tertiary period that it did. But it could have migrated during the mid-Tertiary. Certainly if it did have such restricted geographic extent, then when the placental invasion occurred with the rise of the Panamanian isthmus, the better-developed northern species would have forced Thylacosmilus' extinction by the late Tertiary. If we don't make the restricted-geography assumption, we can't fix the time of Thylacosmilus' extinction. If the species had spread beyond the continent before the rise of the isthmus, it would

[1] Even if we keep (d1) and discard (d2), so that we have only a simple forward persistence axiom, the forward-marching solution is anomalous. It maintains that given the MBA-student is matriculated, we must conclude (defeasibly) that he or she is a first-year student, or more radically, that he or she has just arrived!

have avoided the early extinction met by its marsupial siblings. Symbolically,

Example II.2. (see Figure 3)

SA-restricted@0
Species-intact@0
rise-of-isthmus@1
(1) if SA-restricted@1 and rise-of-isthmus@1 then species-extinct@2
(d1) if SA-restricted@0 then defeasibly SA-restricted@1
(d2) if species-intact@0 then defeasibly species-intact@1
(d3) if species-intact@1 then defeasibly species-intact@2.

Reasoning to not-alive@2 in the gun-firing problem seems to require reasoning to species-extinct@2 in this problem. But there's no reason to suppose extinction over intactness here. We deliberately posed an ambiguous reasoning situation.

Consider Drs. Fiskus and Ehrlich, medical residents who are on-call and must share a single resident's bed. Once in bed, each tends to remain there. But for reasons of propriety, they will not share the bed. To date, they have never shared the on-call bed. One night, Fiskus is in bed at midnight. Ehrlich is tired at midnight and he will be in the bed by the start of the late shift. If Fiskus remains in the bed, then the fact that Drs. Ehrlich and Fiskus have never shared the on-call bed will cease to persist. But it is more plausible that the prospect of sharing the bed with his colleague will cause Dr. Fiskus to get up and get out, preserving their proper professional relationship. Symbolically,

Example II.3. (Figure 3)

never-shared-bed@0
Fiskus-in-bed@0
Ehrlich-in-bed@1
(1) if Ehrlich-in-bed@1 and Fiskus-in-bed@1 then not-never-shared-bed@2
(d2) if Fiskus-in-bed@0 then defeasibly Fiskus-in-bed@1
(d2) if never-shared-bed@0 then defeasibly never-shared-bed@1
(d3) if never-shared-bed@1 then defeasibly never-shared-bed@2.

In each case, we want to represent the problem with the sentences shown. But "temporally forward priority" permits an undesirable inference, unless we add some other information, which would defeat the unwanted inference.

On Hanks' and McDermott's side, there is reason to be suspicious of what has happened. There is an interdependence between representation and inference. As the inference rules and the meaning postulates of the language are changed, so too change the sentences that represent a given situation. We expect that the adoption of new inference rules will change the way

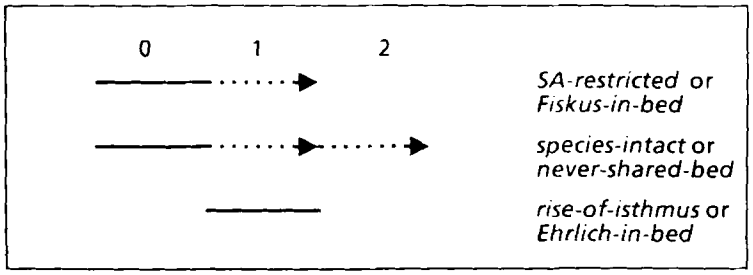

Figure 3.

we represent a given problem. So perhaps the sentences above don't represent the situation as we understand it; perhaps the correct way to represent the situation is to add some sentence that represents the fact that we *don't* want to jump to the Hanks-McDermott conclusion here.

Moreover, non-monotonic inference mechanisms don't guarantee that the conclusions to which they "jump" will be correct. Providing an example in which a non-monotonic rule "guesses" incorrectly does not refute the rule. So, perhaps it is not a refutation of a non-monotonic rule to do what I have done: namely, to show that there are situations in which the rule is a bad rule, irrespective of whether it guesses right or wrong.

Nevertheless, the sentences above so naturally represent the problem described, and the Hanks-McDermott rule mandates a conclusion so egregiously unwarranted, that these cases must be taken as serious challenges to the plausibility of the rule.

So perhaps the Hanks-McDermott intuitions are wrong. The conclusion, not-alive@2, ought not be mandated, at least given the information represented. Nor ought it be mandated that Thylacosmilus was extinct by the late Tertiary period. Nor should be required the sharing of the on-call bed by our reluctant medical residents.

Hanks and McDermott are actually aware of situations in which the represented knowledge has identical form, but the intuition is reversed. Still, they stick to their preferred conclusion and their forward-marching strategy.

> In a richer temporal formalism the criterion [of] chronological minimality might not be the right one. If several years had lapsed between the WAIT and the SHOT, for example, it would be reasonable to assume that the gun was no longer loaded. But chronological minimality does correctly represent our simple notion of persistence: that facts tend to stay true (forever) unless they are "clipped" by a contradictory fact.

It's worth pausing to wonder about this last idea: "facts tend to stay true forever unless clipped...," as if with an inertial veracity. I'll return to it in the end when I speculate on the sources of the odd Hanks-McDermott metaphysics.

III. CAN'T BE HANDLED?

Hanks and McDermott's second belief is that existing systems of non-monotonic reasoning cannot be made to mandate their preferred conclusion. Of course, if the represented facts are not augmented, if the situation is not better qualified, if the sentences stand as they are, then I think it is a virtue of existing systems that they abstain. Otherwise, the system would be vulnerable to the anomalies of reasoning just discussed.

But is there any way to force the not-alive@2 conclusion, possibly by altering the representation of the problem situation?

I can think of two simple ways.[2]

III.1. First Approach

III.1.1. Poole's, Nute's and My Implicit Orderings of Extensions. One way is to add another defeasible rule. It will encode knowledge that appeared to be reflected in the material conditional, but which actually needs to be explicit. With the introduction of this new defeasible rule, we can throw out the material conditional (though we don't have to). Rule (1) was really defeasible anyway. The facts, fired@1 and loaded@1, do not guarantee not-alive@2. What if firing-pin-removed@1? Or finger-in-front-of-hammer@1? The rule comes with a natural set of "unless" conditions, i.e., simple defeaters. And it comes with a set of "even if" conditions, i.e., conditions which, in any combination, do not interfere with the association reflected in the defeasible rule (whether the association is nomological, or causal, or whatever). A good "even if" condition is wearing-after-shave@1. Another is sun-shining-in-Providence@1. Another could be alive@1. This last one is the most interesting.

The antecedent of the defeasible rule could be specialized in any of a number of ways, while still allowing the consequent. In particular, being alive@1 does not interfere with the reported association. So,

(d4) if fired@1 and loaded@1 and alive@1 then defeasibly not-alive@2

can be added to the knowledge base.

---

[2] Henry Kautz has pointed out a standard way to force the Hanks-McDermott intuition. The set

{bird(x) : M fly(x) / fly(x);
penguin(x) : M not-fly(x) / not-fly(x)}

can be changed to

{bird(x) and not-penguin(x) : M fly(x) / fly(x);
penguin(x) : M not-fly(x) / not-fly(x);
bird(x) : M not-penguin(x) / not-penguin(x)}

to force not-fly(Opus) when penguin(Opus).

Similarly, the set of defeasible rules in the original gun example could be altered, though it would leave a morass of unintuitive rules.

Now, I am acquainted with three theories for selecting among multiple extensions: David Poole's (1985), Donald Nute's (1985, 1986), and my own (Loui, 1986a). Each theory says that in the choice between the following lines of reasoning, the latter is superior:

```
{
alive@0;
alive@0 then defeasibly alive@;1
  therefore alive@1
alive@1 then defeasibly alive@2;
  therefore alive@2
}
```

and

```
{
alive@0;
alive@0 then defeasibly alive@1;
  therefore alive@1;
loaded@0;
loaded@0 then defeasibly loaded@1;
  therefore loaded@1;
fired@1;
fired@1 and loaded@1 and alive@1 then defeasibly not-alive@2;
  therefore not-alive@2
}.
```

This superiority is determinable strictly on the basis of syntax. On Poole's account, the latter is "more specific" than the former. Either alive@0 or alive@1 will make the former "applicable" to alive@2, but will not make the latter "applicable" to not-alive@2. Meanwhile, the latter is made applicable only if it is unconditionally known that alive@1 or alive@0, and fired@1, and loaded@0 or loaded@1. That's four possibilities. In each of the four, the former is made applicable, too. Thus, according to Poole, the latter is more specific and hence is preferred.

In Nute's system, (d4) is a "superior non-monotonic rule," because its antecedent, "fired@1 and loaded@1 and alive@1" entails "alive@1," which is the antecedent of its only challenger.

In my system, the latter is superior for a couple of reasons. It has "superior unconditional evidence," i.e., {alive@0} is entailed by {alive@0; loaded@0; fired@1}. It has "superior specificity," i.e., {alive@1 then defeasibly alive@2} is less specific than {fired@1 and loaded@1 and alive@1 then defeasibly not-alive@2}, and this comparison is crucial.

It's hard to imagine a system for selecting among competing defeasible conclusions that would not favor the conclusion with superior evidence, superior specificity, and "equivalent directness" (see Loui, 1986a, for discus-

sion of these properties). It's because specializing the antecedent strengthens the rules in such a way that the rule now dictates what should be done in the multiple extension situation.

III.1.2. Antecedent Inclusiveness. Some would say, justifiedly, that what this amounts to is ordering defaults—encoding the preference information. But it's not ad hoc. It's a genuine part of what is claimed to be known about the "causal" law in this domain.

The reason Hanks and McDermott want to clip alive-ness instead of loaded-ness is that intuitively, they know that alive@1 is one of the assertions that can be in the antecedent of (the defeasible version of) their rule. In short, they hold (d4). What is interesting about the kinds of rules that they have spotlighted is that they have implicit "even if" conditions. Suppose a rule is

(R) if Φ then defeasibly Ψ,

where Φ is a set of properties at some time, $\{\Phi_i@t_0\}$, and Ψ is a set of properties at a later time, $\{\Psi_i@t_1\}$. If (R) is a Hanks-McDermott rule, a "causal" rule, it should be the case that

for any $t^- \leq t_0$, and any Ψ_i s. t. $\Psi_i@t_1 \in \Psi$, if $\{\Psi_i@t^-\} \cup \Phi$ then defeasibly Ψ.

This allows clipping of any property Ψ_i holding at a time t^-.

What is it about alive-ness that allows it to be clipped? It's not just its temporal relation to loaded-ness and the firing. It's the fact that it can appear in the antecedent. Consider properties that hold at subsequent times that can appear in the antecedents. These can be clipped too.

Example III.1. (See Figure 4)
Hiram Sibley was one of George Eastman's well-to-do buddies and was, like Eastman, instrumental in the raising of Rochester into the city in the 20's that many thought would be the next great Eastern metropolis. Let's suppose with an E.L. Doctorowian historical liberty that Sibley had most of his wealth in the stock market just before the great crash. We know that Sibley was rich in the 1910's and again by the 1930's, regardless of what might have happened in 1929 to his net worth. We'd normally reason with backward persistence that

(d1) if rich@3 then defeasibly rich@2.

But we have the rule

(d2) if majority-of-personal-wealth-in-market@1 and market-crash@1 then defeasibly not-rich@2.

And we know the antecedents are true:

majority-of-personal-wealth-in-market@1
stock-market-crash@1
rich@3.

In this case, we will conclude not-rich@2, because what we really know is not just (d2), but also

(d2′) if majority-of-personal-wealth-in-market@1 and market-crash@1 then defeasibly not-rich@2, even if rich@3

or

(d2′′) if majority-of-personal-wealth-in-market@1 and market-crash@1 and rich@3 then defeasibly not-rich@2.

We know that rich@3 can be included in the antecedent without disturbing the association. In short, we know that even in the presence of rich@3, we'd conclude that Mr. Sibley "lost his shirt" in the 1929 crash.

The point is that it is not temporal relatedness, but rather antecedent inclusiveness that is important. This should be very clear if we alter the original gun example. Suppose we know not only that

loaded@0
alive@0
fired@1
(1) if fired@1 and loaded@1 then not-alive@2
(d1) if alive@0 then defeasibly alive@1
(d2) if alive@1 then defeasibly alive@2
(d3) if loaded@0 then defeasibly loaded@1

but also the more "antecedent-specific" or "antecedent-inclusive" rule

(d4) if alive@1 and fired@1 and loaded@1 then defeasibly alive@2.

Suppose also that we have only the original rule (1), and not the enriched version, (1′). Then Poole, Nute, and I now choose the alive@2 extension.

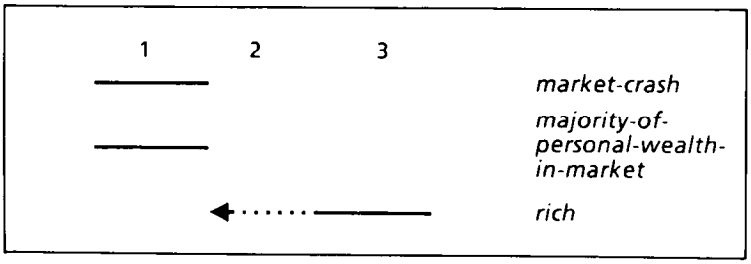

Figure 4.

This is the correct choice given a rule like (d4). Choosing alive@2 just is what (d4) directs us to do, despite loaded-gun-firing.

Consider the example of the extinction of Thylacosmilus and the example of the reluctant medical students, augmented in similar ways.

If we add

(d4) if species-intact@1 and SA-restricted@1 and rise-of-isthmus@1 then defeasibly species-extinct@2

then we are to prefer the conclusion species-extinct@2. The only dissenting chain of reasoning uses the (d3) rule ("if species-intact@1 then defeasibly species-intact@2"). But it has a less specific antecedent, "species-intact@1," which we take to be deferent to the rule with the more specific antecedent.

It's worth wondering whether (d4) should be derivable from (1) ("if SA-restricted@1 and rise-of-isthmus@1 then species-extinct@2"). (1) certainly entails the following rule:

(1′) if species-intact@1 and SA-restricted@1 and rise-of-isthmus@1 then species-extinct@2.

Is (1′) stronger than (d4)? If there is a material connection, an indefeasible connection, then shouldn't there be a "defeasible" connection too? Does "if. . . then defeasibly. . ." also mean "if. . . then (possibly) defeasibly. . ."? Apparently not. At least in Poole's, Nute's, and my systems, the defeasible conditional represents more than a poor man's version of the material connection. Such rules are taken to include implicit directives for choosing among competing defeasible extensions of the theory. Therefore, it must not be the case that material connections express certain connections and defeasible connections express high probability connections, making all material connections defeasible ones as well. Indefeasible connections are not improper or extreme versions of defeasible connections.

Adding

(d4) if Fiskus-in-bed@0 and Ehrlich-in-bed@1 then defeasibly not-Fiskus-in-bed@1

similarly biases the situation in favor of never-shared-bed@2, because

(d1) if Fiskus-in-bed@0 then defeasibly Fiskus-in-bed@1

must now defer to (d4).

III.2. Second Approach

The second simple way that I'd think would accommodate Hanks and McDermott's intuition introduces original-events into the ontology. Considering McDermott's remarks about explanations in his "Critique of Pure

Reason" (1986), it seems that the Hanks-McDermott intuition is based on something like this:

(a) fact: there was a firing event.
(b) if not-alive@2, then there was a dying event, and we know what caused it: the firing event.
(c) if alive@2, then there was still a firing event, and there was also an unloading event.
(d) we don't know what would have caused an unloading event.
(e) so we prefer to reason that there was a dying event rather than an unloading event.

If this is the kind of reasoning desired, then it's appropriate to represent it. Let F, D, D ′ and U be events under our consideration (D, D ′, and U are actually Skolem constants). We have (see Figure 5):

(1) original-event(F)
(2) event-instance(F, 1, firing)
(i.e., F is an instance of a firing event, and F occurred at time 1)
(3) loaded@0
(4) alive@0
(5) (loaded@0 and not-loaded@1) iff event-instance(U, O*, unloading)
(6a) (alive@0 and not-alive@1) iff event-instance(D ′, O*, dying
(6b) (alive@1 and not-alive@2) iff event-instance(D, 1*, dying)
(7a) original-event(U)
(7b) original-event(D ′)
(8) if event-instance(F, 1, firing) and loaded@1 then event-instance(D, 1*, dying) and caused-event(D)
(9) if event-instance(D, 1*, dying) then not-alive@2
(d1) if original-event(x) then defeasibly not-event-instance(x, t, type).

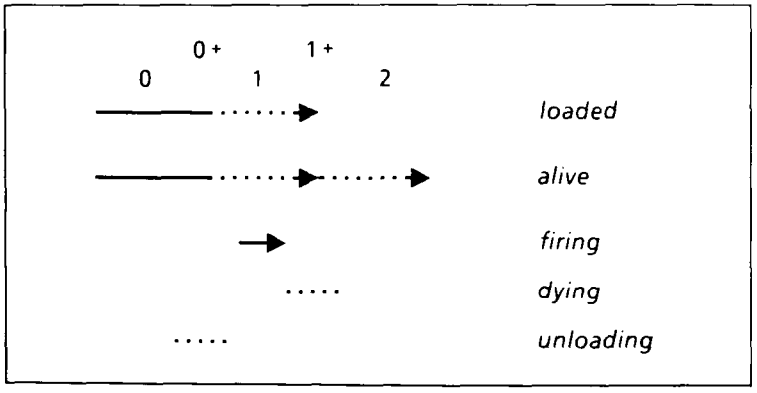

Figure 5.

The crucial bit of knowledge is (7a), which says that we don't know what would have caused an unloading. Of course, it would be difficult to prove this from more primitive principles. But we can just state it here, in all fairness; it's what we know about the problem situation.[3] Now, in effect, we use the defeasible rule (d1) to minimize event-instances for original-events.

(p1) not-event-instance(U, O^+, unloading)	d1, 7a
(p2) not-loaded@0 or loaded@1	p1, 5
(p3) loaded@1	p2, 3
(p4) event-instance(D, 1^+, dying)	p3, 2, 8
(p5) not-alive@2	p4, 9. Q.E.D.

Is this cheating? Perhaps it makes the reasoning "domain-specific," which Hanks and McDermott would like to avoid. It certainly suggests a treatment of events that is not a part of the non-monotonic inference engine. But that's all right. Our representation of time also suggests a treatment of a metaphysical notion that is external to the inferential mechanism.

Structure imposed on problem representation, reflecting one's favorite metaphysics, is not going to be a proper part of the inference engine. Hanks and McDermott may complain at this point that such solutions were obvious to them from the start, but were excluded as bad solutions. If they do, we'll have to take a close look at why they think they're bad solutions. Why do they think it a problem with non-monotonic inference systems that they leave some knowledge representation work to be done?[4]

IV. STRANGE METAPHYSICS

My first point is that from an inference standpoint, the knowledge represented in the original Hanks-McDermott problem—the stuff in syntax—is

---

[3] This information could easily be arrived at defeasibly. Imagine circumscribing explained-events first, then circumscribing original-event-instances, where

original-event-instance(x) iff original-event(x), and $\exists$t, type.event-instance(x, t, type)

explained-event(x) iff not-original-event(x).

The only explained-event would be the dying, D. Thus (7a) and (7b) follow. Then everything proceeds as if (7a) and (7b) had been given.

However, if I had arrived at (7a) and (7b) via (interacting) defeasible rules, the solution would have looked very much like the solution in the first approach.

[4] Even if we agree with the "Yale school" that there is a problem in which not-alive@2 is the correct solution, there is the question of whether that problem is the one that is represented in the syntax. I claim that their syntax equally describes a problem in which not-alive@2 is not a mandatory (defeasible) conclusion. Thus, the syntax—the problem originally represented—is ambiguous at best. More to the point, the problem we would have in mind if we agreed with the Yale school would be improperly stated. Additional represented knowledge is what is required.

insufficient to yield not-alive@2. My second point is that from a knowledge representation standpoint, it's possible to encode the knowledge required to mandate not-alive@2. This is the knowledge that allows Hanks and McDermott to choose not-alive@2 over alive@2. And we are not surprised that this knowledge needs to be represented—to occur in syntax—in order for the intuition to be brought out by the inference engine.

I should note that at least two other authors have shown how to arrive at the Hanks-McDermott intuition with non-monotonic systems. Both involve a slight modification of circumscription (Kautz, 1986; Lifschitz, 1986). Both commit the agent to inferences with temporally forward priority, like Hanks and McDermott.

I don't doubt for a second that common-sense reasoning about competing persistence axioms and causal laws is important. Nor do I doubt that it will occur frequently in an Artificial Intelligence system. But I take issue with the proposed analysis of what seems to be the problem here. Furthermore, I raise my back and curl my tail at the inferential mechanism that Hanks and McDermott, and Shoham too, have recommended.

Why would one even think to draw earlier defeasible conclusions before later ones? It obviously doesn't work for retrospective agents.

I think a clue can be found in the odd Hanks and McDermott statement I quoted earlier: "facts tend to stay true (forever) unless they are 'clipped' by a contradictory fact." This is a strange metaphysics. Usually, we would say that facts about the world, if indeed they are facts, stay true forever. Period. Properties are clipped, not facts. It may be useful to carve up the world into properties and events, in such a way that properties persist except when there are events. We might say that our scientific theories try to minimize spurious clippings, or unexplained events, as a matter of scientific theorizing convention. We might also, as a matter of scientific theorizing convention, propose nomic generalizations (e.g., causal laws) that reflect the past's influence on the future, and not vice versa. None of this explains the inclination to draw forward-marching defeasible conclusions. Beliefs about laws and properties and events are different from the conventions about persistence of properties and occurrences of events.

Facts are facts, and they are timeless. Beliefs, on the other hand, have the kind of inertia that Hanks and McDermott envision. We tend to hold beliefs until we are forced to relinquish them. This is true in both psychological and normative theories about beliefs. It is an epistemological point, not a metaphysical one.

It makes perfect sense to say that beliefs at earlier times should be formed before beliefs at later times. Its truth is analytic. Now consider an agent who is forming beliefs contemporaneously with the occurrence of events. At time $t=0$ believe two things: {alive, loaded}. At $t=1$, the beliefs persist,

and a new belief is added: {fired}.[5] By t = 2, the belief-forming agent is committed to {not-alive}.[6] Could this be what Hanks and McDermott had in mind?

If so, then they have made a conceptual mistake. We should be legislating rational beliefs about temporal relations among properties. That's not the same thing as reproducing the temporal properties of an agent's belief-forming processes. We can state the mistake as a transposition: there is a difference between the temporal evolution of beliefs, and beliefs about temporal evolution. In the case of the gun-firing example, the difference is a matter of life and death.

REFERENCES

Hanks, S., & McDermott, D. (1986). Default reasoning, nonmonotonic logics, and the frame problem. *Proceedings of the American Association for Artificial Intelligence.* Philadelphia, PA. Conference award for best paper.

Harman, G. (1986). *Change in View.* Cambridge, MA: MIT Press.

Kautz, H. (1986). A logic of persistence. *Proceedings of the American Association for Artificial Intelligence.* Philadelphia, PA.

Lifschitz, V. (1986). Pointwise circumscription. *Proceedings of the American Association for Artificial Intelligence.* Philadelphia, PA.

Loui, R. (1986a). *Defeat among arguments: A system of defeasible inference* (Tech. Rep. No. 190). Rochester, NY: University of Rochester, Department of Computer Science. *Computational Intelligence* (to appear).

Loui, R. (1986b). *Real rules of inference: Acceptance and non-monotonicity in AI* (Tech. Rep. No. 191). Rochester, NY: University of Rochester, Department of Computer Science. *Communication and Cognition—AI* (to appear).

McDermott, D. (1986). Critique of pure reason. *Computational Intelligence* (to appear).

Nute, D. (1985). A non-monotonic logic based on conditional logic. Working paper, Advanced Computational Methods Center, University of Georgia, Athens, GA.

Nute, D. (1986). *LDR: A logic for defeasible reasoning* (ACMC Research Rep. No. 01-0013). Athens: University of Georgia, Advanced Computational Methods Center.

Pelavin, R. (1986). *A formal logic that permits planning in temporally rich domains.* Doctoral dissertation, University of Rochester, Department of Computer Science.

Poole, D. (1985). On the comparison of theories: Preferring the most specific explanation. *Proceedings of the International Joint Conference on Artificial Intelligence,* Los Angeles, CA.

Shoham, Y. (1986). Chronological ignorance: Time, nonmonotonicity, necessity, and causal theories. *Proceedings of the Association of Artificial Intelligence,* Philadelphia, PA.

---

[5] Another way to look at it is to say that at t = 0 the agent believes {alive@0, loaded@0}. At t = 1, the agent believes {alive@0, loaded@0, alive@1, loaded@1}. Now the agent comes to believe fired@1 as well. Given the beliefs at t = 1, including a commitment to loaded@1, it is a simple matter to defeat persistence. alive@2 defers to not-alive@2. This is the kind of belief dynamics we'd expect from a coherentist model such as the one Harman (1986) has been advancing.

[6] Actually, there would be belief-revision strategies that do not commit the agent to not-alive, but we can ignore them here.