STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO AIM-153

COMPUTER SCIENCE DEPARTMENT
REPORT NO. CS-242

THE FRAME PROBLEM AND RELATED PROBLEMS ON ARTIFICIAL INTELLIGENCE

BY

PATRICK J. HAYES

NOVEMBER, 1971

COMPUTER SCIENCE DEPARTMENT

STANFORD UNIVERSITY

STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO AIM-153

NOVEMBER 1971

THE FRAME PROBLEM AND RELATED PROBLEMS IN ARTIFICIAL INTELLIGENCE

by

Patrick J. Hayes

Metamathematics Unit, University of Edinburgh
and
Computer Science Dept., Stanford University

ABSTRACT: The frame problem arises in considering the logical structure of a robot's beliefs. It has been known for some years, but only recently has much progress been made. The problem is described and discussed. Various suggested methods for its solution are outlined, and described in a uniform notation. Finally, brief consideration is given to the problem of adjusting a belief system in the face of evidence which contradicts beliefs. It is shown that a variation on the situation notation of (McCarthy and Hayes, 1969) permits an elegant approach, and relates this problem to the frame problem.

The research reported here was supported in part by the Advanced Research Projects Agency of the Office of the Secretary of Defense (SD-183).

Paper presented to the NATO Symposium HUMAN THINKING - COMPUTER TECHNIQUES FOR ITS EVALUATION. St. Maximin, 17-20 August 1971.

Reproduced in the USA. Available from the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.
Price: full size copy, $3.00; microfiche copy, $ .95.

1. INTRODUCTION

We consider some problems which arise in attempting a logical analysis of the structure of a robot's beliefs.

A robot is an intelligent system equipped with sensory capabilities, operating in an environment similar to the everyday world inhabited by human robots.

By belief is meant any piece of information which is explicitly stored in the robot's memory. New beliefs are formed by (at least) two distinct processes: thinking and observation. The former involves operations which are purely internal to the belief system: the latter involves interacting with the world, that is, the external environment and, possibly, other aspects of the robot's own structure.

Beliefs will be represented by statements in a formal logical calculus, called the belief calculus L_b. The process of inferring new assertions from earlier ones by the rules of inference of the calculus will represent thinking. (McCarthy 1959, 1963; McCarthy and Hayes 1969, Green 1969, Hayes 1971).

There are convincing reasons why L_b must include L_c - classical first-order logic. It has often been assumed that a moderately adequate belief logic can be obtained merely by adding axioms to L_c (a first-order -theory); however I believe that it will certainly be necessary to add extra -rules of inference to L_c, and extra richness to handle these extra rules.

One can show that under very general conditions, logical calculi obey the extension property: If $S \vdash p$ and $S \subseteq S'$ then $S' \vdash p$. The importance of this is that if a belief p is added to a set S, then all thinking which was legal before, remains legal, so that the robot need

not check it all out again.

2. TIME AND CHANGE

For him to think about the real world, the robot's beliefs must handle time. This has two distinct but related aspects.

(a) There must be beliefs about time. For example, beliefs about causality.

(b) The robot lives in time: the world changes about him. His beliefs must accommodate in a rational way to this change.

Of these, the first has been very extensively investigated both in A.I. and philosphical logic, while the second has been largely ignored, until very recently: it is more difficult. The first is solely concerned with thinking: the second involves observation.

The standard device for dealing with (a) is the introduction of situation variables (McCarthy 1963, et seq.) or possible worlds (Hintikka 1967; Kripke 1963). Symbols prone to change their denotations with the passage of time are enriched with an extra argument-place which is filled with a term (often a variable) denoting a situation which one can think of intuitively as a time instant, although other readings are possible. In order to make statements about the relationships between situations, and the effects of actions, we also introduce terms denoting events, and the -function R (read: result) which takes events and situations into new situations. Intuitively, 'R(e,s)' denotes the situation which results when the event e happens in the situation s. By 'event' we mean a change in the world: "his switching on the light", "the explosion", "the death of Caesar". This is a minor technical simplification of the notation and terminology used in (McCarthy and Hayes 1969) and(Hayes 1971).

Notice that all the machinery is defined within L_c. The situation calculus is a first-order theory.

Using situations, fairly useful axiomatizations can be obtained for a number of simple problems involving sequences of actions and events in fairly complicated worlds. (Green 1969), (McCarthy and Hayes 1969).,

3. THE FRAME PROBLEM

Given a certain description of a situation s - that is, a collection of statements of the form $\phi[\![s]\!]$, where the fancy brackets mean that every situation symbol in ϕ is an occurrence of 's' - we want to be able to infer as much as possible about $R(e,s)$. Of course, what we can infer will depend upon the properties of e. Thus we require assertions of the form:

$$\phi_1[\![s]\!] \;\&\; \psi(e) \supset \phi_2[\![R(e,s)]\!] \qquad (1)$$

Such an assertion will be called a law of motion. The frame problem can be briefly stated as the problem of finding adequate collections of laws of motion.

Notice how easily human thinking seems to be able to handle such inferences. Suppose I am describing to a child how to build towers of bricks. I say "You can put the brick on top of this one onto some other one, if that one has not got anything else on it." The child knows that the other blocks will stay put during the move. But if I write the corresponding law of motion:

$$(on(b_1,b2,s) \;\&\; \forall z.\; \neg\, on(z,b_3,s)) \supset on(b_1,b_3,R(move(b_2,b_3),s)) \qquad (2)$$

then nothing follows concerning the other blocks. What assertions could we write down which would capture the knowledge that the child has about

the world?

One does not want to be obliged to give a law of motion for every aspect of the new situation. For instance, one feels that it is prolix to have a law of motion to the effect that if a block is not move'd, then it stays where it is. And yet such laws - instances of (1) in which $\phi_1 = \phi_2$ - are necessary in first-order axiomatizations. They are called frame axioms. Their only function is to allow the robot to infer that an event does not affect an assertion. Such inferences are necessary: but one feels that they should follow from more general considerations than a case-by-case listing of axioms, especially as the number of frame axioms increases rapidly with the complexity of the problem. Raphael (1971) describes the difficulty thoroughly.

This phenomenon is to be expected. Logically, s and R(e,s) are simply different entities. There is no a priori justification for inferring any properties of R(e,s) from those of s. If it were usually the case that events made widespread and drastic alterations to the world (explosions, the Second Coming, etc.), then we could hardly expect anything better than the use of frame axioms to describe in detail, for each event, exactly what changes it brings about. Our expectation of a more general solution is based on the fact that the world is, fortunately for robots, fairly stable. Most events - especially those which are likely to be considered in planning - make only small local changes in the world, and are not expected to touch off long chains of cause and effect.

4. FRAME RULES

We introduce some formalism in order to unify the subsequent discussions. Any general solution to the frame problem will be a method

for allowing us to transfer properties from a situation s to its successor $R(e,s)$; and we expect such a license to be sensitive to the form of the assertion, to what is known about the event e, and possibly to other facts.

Consider the rule scheme FR:

$$\chi,\ \phi[\![s]\!],\ \psi(e) \vdash \phi[\![R(e,s)]\!]$$

provided $\mathcal{H}(e, \phi, \psi)$. (FR)

where $\mathcal{H}$ is some condition on e, ϕ and ψ, expressed of course in the metalanguage. We will call such a rule a frame rule. The hope is that frame rules can be used to give a general mechanism for replacing the frame axioms, and also admit an efficient implementation, avoiding the search and relevancy problems which plague systems using axioms (Green 1969, Raphael 1971).

One must, when considering a frame rule, be cautious that it does not allow contradictions to be generated. Any addition of an inference rule to L_c, especially if not accompanied by extra syntax, brings the risk of inconsistency, and will, in any case, have dramatic effects on the metatheory of the calculus. For instance, the deduction theorem fails. Thus a careful investigation of each case is needed. In some cases, a frame rule has a sufficiently simple $\mathcal{H}$ condition that it may be replaced by an axiom scheme, resulting in a more powerful logic in which the deduction theorem holds. This usually makes the metatheory easier and implementation more difficult.

5. SOME PARTIAL SOLUTIONS USING FRAME RULES

The literature contains at least four suggestions for handling the

problem which are describable by frame rules. In each case we need some extra syntactic machinery.

(1) Frames (McCarthy and Hayes 1969)

One assumes a finite number of monadic second-order predicates P_i. If $\vdash P_i(h)$ for a nonlogical symbol h (predicate, function or individual constant) then we say that h is in the i^{th} block of the frame. The frame rule is

$$P_{i_1}(h_1), \ldots, P_{i_n}(h_n), \phi[\![s]\!], P_j(e) \vdash \phi[\![R(e,s)]\!] \qquad (6)$$

where $h_1, \ldots, h_n$ are all the nonlogical symbols which occur crucially in ϕ, and $i_k \neq j$, $1 \leq k \leq n$.

Here crucial is some syntactic relation between h and ϕ; different relations give different logics, with a stronger or weaker frame rule.

(2) Causal Connection (Hayes 1971)

We assume that there is a 3-place predicate $\rightarrow(x,y,s)$ (read: x is connected to y in situation s) which has the intuitive meaning that if x is not connected to y, then any change to y does not affect x. It seems reasonable that $\rightarrow$ should be a partial ordering on its first two arguments (reflexive and transitive). The frame rule is:

$$\phi[\![s]\!], \neg\rightarrow(h_1,e,s), \ldots, \neg\rightarrow(h_n,e,s) \vdash \phi[\![R(e,s)]\!] \qquad (7)$$

where (i) ϕ is an atom or the negation of an atom

(ii) $h_1, \ldots, h_n$ are all the terms which occur crucially in ϕ.

If we insisted only that $\rightarrow(h_i,e,s)$ is not provable (rather than $\neg\rightarrow(h_i,e,s)$ is provable) then the rule is much stronger but no longer obeys the extension property. This is analogous to PLANNER'S method below.

(3) MICRO-PLANNER (Sussman and Winograd 1969)

The problem-solving language MICRO-PLANNER uses a subset of

predicate calculus enriched with notations which control the system's search for proofs. We will ignore the latter aspect for the present and describe the underlying formalism. Its chief peculiarity is that it has no negation, and is therefore not troubled by the need for consistency.

Following MICRO-PLANNER we introduce the new unary propositional connective therase. Intuitively, therase ϕ will mean that ϕ is 'erased'. We also introduce the notion of a transition: an expression $\langle e{:}\phi_1,\ldots,\phi_n\rangle$. This means intuitively 'erase $\phi_1,\ldots,\phi_n$ in passing from s to R(e,s)'. The frame rule is:

$$\chi,\ \phi[\![s]\!],\ \langle e{:}\ \phi_1,\ldots,\phi_n\rangle \vdash \phi[\![R(e,s)]\!] \qquad (8)$$

where (i) ϕ is an atom;

(ii) ϕ contains no variables (other than s);

(iii) χ, therase $\phi_1,\ldots,$therase $\phi_n \not\vdash$ therase $\phi[\![s]\!]$

Notice the negated inference in (iii).

(4) STRIPS (Fikes and Nilsson 1971)

The problem-solving system STRIPS uses the full predicate calculus enriched with special notations ('operator descriptions') describing events, and ways of declaring certain predicates to be primitive. We can use transitions to describe this also. The frame rule is:

$$\phi[\![s]\!],\ \langle e{:}\ \phi_1,\ldots,\phi_n\rangle \vdash \phi[\![R(e,s)]\!] \qquad (9)$$

where (i) ϕ is an atom or the negation of an atom

(ii) ϕ contains no variables (other than s)

(iii) the predicate symbol in ϕ is primitive

(iv) $\phi[\![s]\!]$ is not an instance of any ϕ_i, $1 \le i \le n$.

Notice the similarity to (8). Primitive can be axiomated by the use of a

monadic second-order predicate, as in (1) above.

These four rules have widely divergent logical properties. Rule (6) is replaceable by an axiom scheme, and is thus rather elementary. It is also very easy to implement efficiently (theorem-proving cognoscenti may be worried by the higher-order expressions, but these are harmless since they contain no variables). Variations are possible: e.g., we might have disjointness axioms for the P_i and require $\neg P_j(h_k)$ rather than $P_{i_k}(h_k)$: this would be closely similar to a special case of (7).

Retaining consistency in the presence of (6) requires in nontrivial problems that the P_i classification be rather coarse. (For instance, no change in position ever affects the color of things, so predicates of location could be classed apart from predicates of color.) Thus frames, although useful, do not completely solve the problem.

Rule (7) is also replaceable by an axiom scheme, and the restriction to literals can be eliminated, with some resultant complication in the rule. Also, there is a corresponding model theory and a completeness result (Hayes 1971), so that one can gain an intuition of what (7) means. Retaining consistency with (7) requires some care in making logical definitions.

Rules (8) and (9) have a different character. Notice that (9) is almost a special case of (8): that in which therase ϕ ⊢— therase ψ iff ψ is not primitive or ψ is an instance of ϕ. The importance of this is that instantiation, and probably primitiveness also, are decidable, and conditions (iii) and (iv) in (9) are effectively determined solely by examining the transition, whereas condition (iii) in (8) is in general not decidable and in any case requires an examination of all of χ: in applications, the whole set of beliefs. MICRO-PLANNER uses its ability to

control the theorem-proving process to partly compensate for both of these problems, but with a more expressive language they would become harder to handle. Notice also that (8) does not satisfy the extension property, while (9) does, provided we allow at most one transition to be unconditionally asserted for each event.

Maintaining 'consistency' with (8) is a matter of the axiom-writer's art. There seem to be no general guidelines. Maintaining consistency with (9) seems to be largely a matter of judicious choice of primitive vocabulary. There is no articulated model theory underlying (8) or (9). They are regarded more as syntactic tools - analogous to evaluation rules for a high level programming language - than as descriptive assertions.

6. A (VERY) SIMPLE EXAMPLE: TOY BRICKS.

A1. $\neg$ above (x,x,s)

A2. x = Table $\vee$ above (x,Table,s)

A3. above(x,y,s) $\equiv .$ on(x,y,s) $\vee \exists$ z.on(z,y,s)& above(x,z,s)

A4. free(x,s) $\equiv .$ $\forall$y $\neg$ on(y,x,s)

To enable activity to occur we will have events move(x,y): the brick x is put on top of the brick y. Laws of motion we might consider include:

A5. free(x,s)& x $\neq$ y. $\supset$ on(x,y,R(move(x,y),s))

... free(x,s)& w $\neq$ x & on(w,z,s). $\supset$ on(w,z,R(move(x,y),s))

A7. free(x,s)& w $\neq$ x & above(w,z,s). $\supset$ above (w,z,R(move(x,y),s))

A8. free(x,s)& w $\neq$ y & free(w,s). $\supset$ free(w,R(move(x,y),s))

Of these, A6-A8 are frame axioms. (In fact, A7 and A8 are redundant, since they can, with some difficulty, be derived from A6 and A3, A4 respectively.)

A5 assumes somewhat idealistically that there is always enough space on γ to put a new brick.

Rule (6) cannot be used in any intuitively satisfactory way to replace A6-A8.

Rule (7) can be used. We need only to specify when bricks are connected to events:

A9. $\rightarrow(x,\text{move}(y,z),s) \equiv . \; x = y \vee \text{above}(x,y,s)$

Using A9 and A3, A4, it is not hard to show that

$$\text{free}(x,s) \,\&\, w \neq x \,\&\, \text{on}(w,z,s) . \supset . \neg \rightarrow (w,\text{move}(x,y),s) \,\&\, \neg \rightarrow (z,\text{move}(x,y),s)$$

and thus, we can infer $\text{on}(w,z,R(\text{move}(x,y),s))$ by rule (7). A7 and A8 are similar but simpler. (One should remark also that A4 is an example of an illegal definition, in the presence of (7), since it suppresses a variable which the rule needs to be aware of. It is easy to fix this up in various ways.)

Rule (8) can also be used, but we must ensure that therase does a sufficiently thorough job. Various approaches are possible. The following seems to be most in the spirit of MICRO-PLANNER. In its terms, on and above statements will be in the data-base, but free statements will not. The necessary axioms will be:

A10. therase $\text{free}(x,s)$

A11. therase $\text{on}(x,y,s) \,\&\, \text{above}(y,z,s) \supset$ therase $\text{above}(x,z,s)$

A12. $\text{free}(x,s) \supset \langle \text{move}(x,y) : \text{on}(x,z,s) \rangle$

To infer statements $\text{free}(x,R(e,s))$, we must first generate enough $\text{on}(x,y,R(e,s))$ statements by rule (8), and then use A4, since by A10, rule (8) never makes such an inference directly. (We could omit A10 and replace by A12 by:

A13. $\text{free}(x,s) \; 3 \; (\text{move}(x,y): \text{on}(x,z,s), \text{free}(y,s)\rangle .$

This would, in MICRO-PLANNER terms, be a decision to keep free assertions in the data base.)

Notice that MICRO-PLANNER has no negation and hence no need to therase such assertions as $\neg$ on(x,y,s). If it had negation we would replace A12 by

A14. free(x,s) $\supset$ ⟨move(x,y): on(x,z,s) , $\neg$ on(x,y,s)⟩

and add

A15. **therase** $\neg$ on(x,y,s) & above(y,z,s) $\supset$ therase $\neg$ above(x,z,s)

Notice the close relations between A3, A11 and A15.

Rule (9) can be used similarly to (8), but we are no longer able to use axioms such as A11 and A15. The solution which seems closest in spirit to STRIPS is to declare that on is primitive but that above and free are not, and then simply use A14. The 'world model' (Fikes and Nilsson 1971) would then consist of a collection of atoms on(a,b), or their negations, and the system would rederive above and free assertions when needed. This is very similar to MICRO-PLANNER'S 'data-base', and we could have used rule (8) in an exactly similar fashion.

7. IMPLEMENTING FRAME RULES

Some ingenuity with list structures enables one to store assertions in such a way that

(i) Given s, one can easily find all assertions $\phi[s]$;

(ii) Each symbol denoting a situation is stored only once;

(iii) The relationships between s and R(e,s), etc., are stored efficiently and are easily retrieved;

(iv) To apply a frame rule to s, one need only;

(a) create a new cell pointing to s;

(b) move two pointers;

(c) check each $\phi[\![s]\!]$ for condition $\mathcal{N}$: if it holds, move one pointer.

In the case of a rule like (8) or the variation to (7), where $\mathcal{N}$ is a negative condition ($\vdash\!\!\!/$), we need only examine those $\phi[\![s]\!]$ for which the condition fails resulting in greater savings.

Space does not permit a description of the method, but MICRO-PLANNER and STRIPS use related ideas. (The authors of these systems seem to confuse to some extent their particular implementations with the logical description of the frame rules, even to the extent of claiming that a logical description is impossible.)

8. CONSISTENCY AND COUNTERFACTUALS

Frame rules can be efficiently implemented and, in their various ways, allow the replacement of frame axioms by more systematic machinery. But there is a constant danger, in constructing larger axiomatizations, of introducing inconsistency. An alternative approach avoids this by transferring properties ϕ from s to R(e,s) as long as it is consistent to do so, rather than according to some fixed-in-advance rule.

Suppose we have a set χ of general laws which are to hold of every situation , and a description of - a set of assertions about - the situation s: $\{\phi_1[\![s]\!], \ldots, \phi_n[\![s]\!]\}$. Using laws of motion we will directly infer certain properties $\psi_1, \ldots, \psi_m$ of R(e,s): the set of these constitutes a partial description of R(e,s). To compute a more adequate one, we add assertions $\phi_i[\![R(e,s)]\!]$ in some order, checking at each stage for consistency with χ; if a $\phi_i[\![R(e,s)]\!]$ makes the set inconsistent, it is rejected. This continues until no more ϕ_i can be added. In this way we compute a maximal consistent subset (MCS) of the inconsistent set

$$\chi \cup \{\psi_1, \ldots, \psi_m, \phi_1[\![R(e,s)]\!], \ldots, \phi_n[\![R(e,s)]\!]\}.$$

There are two big problems. One, consistency is not a decidable or even semi-decidable property. Thus for practicality one has to accept a large restriction on the expressive power of the language. Two, there are in general many different MCS's of an inconsistent set, and so we must have ways of choosing an appropriate one. In terms of the procedure outlines above, we need a good ordering on the ϕ_i.

This procedure is closely similar to one described by Rescher (1964) to provide an analysis of counterfactual reasonings ('If I had struck this match yesterday, it would have lit', when in fact I didn't.) Rescher is aware of the first problem but gives no solution. His major contribution is to the second problem, which he solves by the use of modal categories: a hierarchical classification of assertions into grades of law-like-ness. One never adds $\phi_i[\ R(e,s)\ ]$ unless every ϕ_{ij} with a lower classification has already been tested. This machinery is especially interesting as in (Simon and Rescher 1966) it is linked to Simon's theory of causality (Simon 1953). One puts ϕ_i in a lower category than ϕ_j just in case ϕ_i causes ϕ_j (or $\neg\phi_j$), more or less. Space does not permit a complete description of this interesting material which is fully covered in the references cited. In spite of its appeal, the first problem is still unsolved.

In unpublished work at Stanford, Jack Buchanan has independently worked out another version of the procedure. The first problem is handled by accepting a drastic restriction on the language. Every ϕ_i is an atom or the negation of an atom - c.f. frame rules 7, 8 and 9 - and, more seriously, χ contains only assertions of the form $t_1 \neq t_2$ or of the

form $P(t_1,...,t,...,t_n)$ and $P(t_1,...,u,...,t_n) \supset t = u$. Under these constraints, consistency is decidable and can even be computed quite efficiently. Moreover, MCS's are unique, so the second problem evaporates. However, it is not clear whether nontrivial problems can be reasonably stated in such a restricted vocabulary.

9. CONCLUSIONS

In the long run, I believe that a mixture of frame rules and consistency-based methods will be required for nontrivial problems, corresponding respectively to the 'strategic' and 'tactical' aspects of computing descriptions of new situations. In the short term we need to know more about the properties of both procedures.

One outstanding defect of present approaches is the lack of a clear model theory. Formal systems for handling the frame problem are beginning to proliferate, but a clear semantic theory is far from sight,. Even to begin such a project would seem to require deep insight into our presystematic intuitions about the physical world.

10. OBSERVATIONS AND THE QUALIFICATION PROBLEM

We have so far been entirely concerned with thinking. The situation calculus is a belief calculus for beliefs about time. Observations - interactions with the real world - introduce new problems. We must now consider the second aspect of time (2. (b) above).

Almost any general belief about the result of his own actions may be contradicted by the robot's observations. He may conclude that he can drive to the airport only to find a flat tire. A human immediately says "Ah, now I cannot go". Simply adding a new belief ('the tire is flat') renders an earlier conclusion false, though it was a valid conclusion from

the earlier set of beliefs, all of which are still present. Thus we do not assume that the robot had concluded 'If my tires are OK, then I can get to the airport' since there are no end of different things which might go wrong, and he cannot be expected to hedge his conclusions round with thousands of qualifications. (McCarthy and Hayes 1969).

Clearly this implies that the belief logic does not obey the extension property for observations: to expect otherwise would be to hope for omnipotence. However, we are little nearer any positive ideas for handling the inferences correctly.

John McCarthy recently pointed out to me that MICRO-PLANNER has a facility (called THNOT) which apparently solves the problem nicely. I will translate this into a slightly different notation.

We introduce a new unary propositional connective proved, which is supposed to mean 'can be proved from the current collection of beliefs'. Then we can write axioms like the following:

A16. flat(tire) $\supset$ kaput (car)

A17. $\neg$ proved kaput (car) $\supset$ at (robot, airport, R(drive(airport)+))

from which at(robot,airport,...) should be concluded until we add:

A18. flat (tire)

at which point the $\neg$ proved...becomes false. ($\neg$ proved is PLANNER's THNOT).

To make this work we could try the following rules of inference:

$\phi \vdash$ proved ϕ (P1)

$\chi \vdash \neg$ proved ϕ (P2)

where $\chi \nvdash \phi$

P2 fails the extension property, as expected. (It also has the difficulties of effectiveness which worry frame rule (8), but we will ignore these.)

Unforntunately, P1 and P2 are inconsistent. Suppose $\chi \nvdash \phi$, but

that ϕ is consistent with X. Then by P2, $\neg$ proved ϕ. If we now add ϕ (an observation: the flat tire), then by P1 proved ϕ: an overt contradiction. MICRO-PLANNER avoids this by denying P1 and treating 'ϕ & $\neg$ proved ϕ' as consistent. But this is a council of despair, since it clearly is not, according to the intuitive meanings.

The logical answer is to somehow make proved refer to the set X of antecedents. The direct approach to this requires extremely cumbersome notation and a very strong logic which partly contains its own methatheory, thus coming close to Godel inconsistency. Fortunately we do not need to describe sets X of assertions, but only to refer to them, and this can be done with a very weak notation, similar to situation variables.

Assume that every belief is decorated with a constant symbol called the index: we will write it as a superscript. Indices denote the robot's internal belief states just as situation terms denote external situations. Observations are analogous to events. Assertions proved ϕ now have an extra index which identifies the state of belief at the time the inference was tested. The above rules of inference become:

$$\phi^s \vdash \text{proved}^s\ \phi^s \qquad \text{P1'}$$

$$X \vdash \neg\ \text{proved}^s\ \phi^s \qquad \text{P2'}$$

where $X \not\vdash \phi^s$ and every member of X has index s.

In applications we now insist that

(i) In applying P2', X contains all beliefs with index s;

(ii) Whenever an observation is added to the beliefs, every index s is replaced by a new one s', except those on proved assertions.

This is just enough to avoid inconsistency; it clearly does not

involve any Godel-ish difficulties; and (ii) can be very efficiently implemented by frame rule methods (Section 7 above). Indeed, more complex versions of (ii) which allow for direct contradiction between beliefs and observations can be similarly implemented.

'The logic of these indices is trivial, but extensions have some interest. For instance, if we identify indices with situation terms, then expressions of the form $\phi[s]^s$ become legal, with the intuitive meaning 'ϕ is true now'.

Seen this way, the qualification problem is closely linked with the frame problem, and one expects progress in either area to help with the other.

11. ACKNOWLEDGEMENTS

Many people have helped me by giving their time in conversations. I would like to thank Jack Buchanan, Richard Fikes, John McCarthy, Malcolm Newey, Nils Nilsson, Johns Rulifson, Richard Waldinger and Richard Weyrauch. Most of all, I thank my wife, Jackie, for improving my English, controlling my verbosity, and typing innumerable drafts of the manuscript.

REFERENCES

[1] Fikes, R. and Nilsson, N. (1971). "STRIPS: A New Approach to the Application of Theorem-Proving to Problem-Solving." Proceedings of the 2nd International Joint Conference on Artificial Intelligence, Imperial College, London.

[2] Green, C.C., (1969). "Theorem-Proving by Resolution as a Basis for Question-Answering Systems." Machine Intelligence 4. (eds. Meltzer, B. and Michie, D.). Edinburgh University Press.

[3] Hayes, P.J. (1971). "A Logic of Actions." Machine Intelligence 6. (eds. Meltzer, B. and Michie, D.). Edinburgh University Press.

[4] Hintikka, J. (1967). "A Program and a Set of Concepts for Philosophical Logic." The Monist, Vol. 51, p. 67-72.

[5] Kripke, S. (1963). "Semantic Analysis of Modal Logic I." Zeitschrift fur Math. Logik und Grundlagen der Mathematik, Vol. 9, p. 67-96.

[6] McCarthy, J. (1959). "Programs With Common Sense." Mechanization of Thought Processes. Vol. 1. London: Her Majesty's Stationery Office. Reprinted in Semantic Information Processing (ed. Minsky, M.) 1970, MIT Press.

[7] McCarthy, J. (1963). Situations, Actions and Causal Laws. Stanford A.I. Project, Memo 2. Reprinted in Semantic Information Processing, (ed. Minsky, M.) 1970, MIT Press.

[8] McCarthy, J. and Hayes, P.J. (1967). "Some Philosophical Problems From the Standpoint of Artificial Intelligence", Machine Intelligence 4, (eds. Meltzer, B. and Michie, D.), Edinburgh University Press.

[9] Raphael, B. (1971). "The Frame Problem in Problem-solving Systems." Proceedings of the AS1 on Artificial Intelligence and Heuristic Programming, Edinburgh University Press.

[10] Simon, H.A. (1953). "Causal Ordering and Identifiability". Studies in Econometric Method (eds. Hood, W.C. and Koopmans, T.C.). John Wiley.

[11] Simon, H.A. and Rescher, N. (1966). "Cause and Counterfactual". Philosophy of Science, Vol. 33, Bruges, Belgium.

[12] Sussman, G. and Winograd, T. (1969). Micro-Planner Reference Manual. Internal Memorandum, Artificial Intelligence Group, M.I.T.