28

Robotologic

P. J. Hayes
Metamathematics Unit
University of Edinburgh

INTRODUCTION

A robot, in order to act intelligently, must be able to reason from facts which its sensors detect to conclusions which govern its actions. This reasoning process is so central to human intelligence that it seems immediately relevant to the problems of robot design to consider its properties, how it might be analysed and imitated.

Strawson (1959) defines *descriptive metaphysics* as the study of the

'massive central core in human thinking which has no history – or none recorded in histories of thought; there are categories and concepts which, in their most fundamental character, change not at all. Obviously these are not the specialities of the most refined thinking. They are the commonplaces of the least refined thinking; and are yet the indispensable core of the conceptual equipment of the most sophisticated human beings. It is with these, their inter-connexions, and the structure that they form, that a descriptive metaphysics will be primarily concerned.'

This paper may be looked upon as an attempt at the beginnings of a *synthetic* metaphysics; an attempt at an initial description and analysis of some aspects of the choice of a language and construction of an axiomatic theory suitable for the robot's central reasoning agency.

The reader will find that many problems are described, but that almost no solutions are offered. To some extent, this is inevitable at the present time; but in any case I feel that an analysis of what the problems *are*, should precede attempts to solve them.

In the first part of the paper, some general considerations on the construction of a suitable theory are outlined; in the second part, the choice of a formal language is discussed; and in the third part is an exposition of some results and possibilities for dealing with *time*.

My approach to robotology is firmly in the tradition of the advice-taker and SRI robot projects (McCarthy 1968, McCarthy and Hayes 1969, Green 1969), and I owe a large debt of gratitude to the workers at Stanford and

SRI, especially to Professor John McCarthy and Dr Cordell Green, for their advice and encouragement, and for creating an ambience in which I feel at home.

1. CONSTRUCTING A THEORY

In constructing a theory one consideration stands out as paramount. The robot must use its reasoning abilities to decide what to do – to decide upon courses of action. It must be possible for it to reach an appropriate decision in a fairly short time: in particular, sufficiently short that the environment has not changed so much since the last observation that the action is no longer appropriate. This need for 'almost-decidability' is the first and most severe requirement which the theory must satisfy.

It is possible to render any theory decidable in a trivial way by invoking a time cutoff on reasonings and having a default mechanism for deciding the values of any expressions still not decided. Such a mechanism will undoubtedly have to be present as a safety precaution. But if the basic theory is so unmanageable that the default mechanism is invoked frequently, then the robot will be making inappropriate responses more often than not, since of course the standardized response is likely not to be appropriate for more than a very small class of situations. It will probably be analogous to the 'freeze!' response which mammals give to threatening and uncontrollable environments. There does not seem to be any way of avoiding the conclusion that the basic theory must admit an efficient theorem-proving procedure which is close to being a decision procedure.

This conclusion is also not escaped by the idea of using special-purpose routines for certain classes of situations, since then the *decision to enter a certain routine* must be made by computations within the central theory. The necessity of avoiding long delays in decision making is still apparent.

In order to achieve this computational efficiency a compromise is necessary between two conflicting requirements. One is that the *expressive power* of the theory should be as great as possible, so that proofs are not of exorbitant length. The other is that the *search space for proofs* should be as *small* as possible.

Both of these are clearly related in a direct way to efficiency. If the proof desired is excessively long, then there will not be time to generate it even if the search algorithm is perfect. This is dramatically illustrated by McCarthy's first-order axiomatization of the mutilated checkboard problem (McCarthy 1964). If the search space is large, filled with proofs of irrelevant theorems, then even short proofs will be hard to find: there will be no efficient search algorithm.

Unfortunately these two requirements are in direct conflict. A rich, expressive theory is, by definition, one in which many facts can be phrased and in which many theorems can be proved. Thus the search space resulting will be correspondingly rich and difficult to search. If restrictive conditions are imposed upon the form of allowable proofs, then the space is cut down but the theory is weakened and proofs are lengthened. This is common lore in theorem-proving: all the restrictive strategies (hyper-resolution (Robinson 1965), set of support (Wos, Carson and Robinson 1965) for instance) have this effect of at once reducing the size of the search space and in general increasing the length of proofs. The need for compromise has been noted in the set-of-support context.

The lack of expressive power of first-order logic has been noted on several occasions, notably by McCarthy in the reference cited above. The use of richer languages, for instance higher-order logic, has been put forward as a desirable goal to overcome the resulting difficulties. But such rich languages suffer from the other defect above, namely, the extremely large search space thus created for a theorem-proving program.

Consider for instance the question of instantiation, the logical deduction of an instance of an expression from the expression itself. In first-order logic the only expressions which need be substituted for variables during a search are the members of the Herbrand universe, which has the simple structure of a free algebra on a small number of generators. Even so, a direct search through all Herbrand instances yields a hopelessly inadequate theorem-proving algorithm. The giant step in theorem-proving research was the construction of a method whereby the process of instantiation could be controlled so that only those instances were automatically selected which could potentially be used in a proof. This is what the well-known unification algorithm achieves (Robinson 1965, Prawitz 1960). It is now clear that this is effectively *all* the control that *can* be had over instantiation, if completeness is to be preserved. One does not therefore hope for a further major breakthrough; this feeling is borne out by the lack of significant improvements in theorem-proving efficiency since the first Resolution programs, in spite of the greatly expanded research effort. What progress has been made has been due to use of more sophisticated searching and retrieval algorithms within the context of the unification and matching process.

When one comes to higher-order logic the instantiation process is vastly more difficult to control. If most of the power of the language is not to be sacrificed, then a potential instance of a higher-order variable is any lambda-expression of the appropriate type. The set of all such expressions has a far more complex structure than the first-order Herbrand universe. Moreover, no method is known for controlling the instantiation process in higher-order logic: indeed, it is known that no matching algorithm exists (Gould 1966). Without such a device the search is hopeless. Consider a particularly simple example, namely, the induction axiom

$$\forall P(P(0) \wedge (\forall x P(x) \supset P(x+1)) \supset \forall x P(x)). \qquad (1)$$

Searching for instances of (1) involves enumerating *all properties of numbers* and trying them in some order: clearly a hopeless task. It is precisely this job of selecting the appropriate instances of the induction axiom which is often held to constitute the 'creative' element of a proof: herein lies the 'concept-formation'.

This is one illustration of the antagonism noted above between expressive power and search efficiency, and there are others. I do not feel that there is any way of avoiding the dilemma, but there have been some suggestions. The most notable ones are that theorem-provers will continue to improve in efficiency at the same dramatic rate as they did during the last decade – I have indicated above why I think this is a vain hope – and that the problem is relieved by using heuristics. Thus one is supposed to use a rich, expressive language and control the resulting gigantic search by the rigorous use of powerful pruning heuristics. But where are the heuristics to come from? This can only have two answers – either we are really reduced to some weaker theory again or else we are cheating by invoking an 'intelligent subroutine'. The argument has some validity in applying theorem-proving to highly specialized problem domains, where powerful heuristics are available, and also in interactive applications where the human is the intelligent subroutine (Guard *et al.* 1969, Allen and Luckham 1970), but for the robot, generality is all-important and powerful – problem dependent – heuristics just will not be available.

This is not to say that *no* heuristic power can be used: syntactic heuristics of the kind described by Quinlan and Hunt (1968) of course will be useful. But these are essentially *weak*, since they have no information available to them other than the syntactic form of the expressions themselves. Powerful heuristics are notoriously application-dependent.

1.1 Ontology

These rather general considerations can be applied to a concrete problem in theory construction which, although of fundamental importance, has not received any explicit attention in the literature on robots. This is the question of what ontology is to act as the foundation of the theory; or, to put it in less philosophical terms, the question of what entities the robot is to believe exist. Another way of phrasing it in the (present) context of formalized languages is as the problem of what entities are to be regarded as being values of bound variables (this is Quine's dictum: 'to be, is to be the value of a bound variable') and (although there are some technical complexities here) of individual constants. This rendering of the statement of the problem makes clear the relationship between ontological commitment and the sort of formal – and hence computational – complexity of theories discussed above.

The question of ontology has been much discussed by modern philosophers. I will outline two distinct approaches to it, and discuss them with a view to their use by a robot. Of course there are others; and my necessarily brief remarks will not do justice to the arguments used by the philosophers in question; and their motivations in constructing these ontological systems are quite different; nevertheless some useful points seem to emerge.

The first view may be caricatured as *set-theoretic platonism.* It is the view that *sets* exist. By Quine's dictum, anyone who advocates the inclusion of set theory in his theory must admit to the view that sets exist: and set theory is widely held to be at the basis of all of mathematics.

Now, set theory is the example *par excellence* of a rich, expressive theory. It is also one of the formally most intractable theories ever devised, as the huge literature concerned with elucidating its structure testifies. It seems quite impossible that a robot could be made to use the whole of axiomatic set theory in an economical way.

All the objections detailed above to higher-order logic apply with greater force to set theory. In particular, the instantiation problem reappears in the guise of the axiom schema of comprehension which allows us to assume the existence of a set corresponding to any* well-formed formula of the language with one free variable. Many of the axioms simply give one a licence to construct sets: thus the axioms of union, pairing and power-set all have this character. Axiomatic set theory provides a rich field *within which* mathematical creativity may be exercised; but it is far too large a field to *search.* It is, of course, not intended to be otherwise.

It is true (cf. below) that certain features of set theory are attractive and that certain weaker theories which include some sets in their ontologies (weak second-order theories for instance) may be useful: but that is another question.

An opposite extreme from platonism is *nominalism,* of which one of the most distinguished exponents is Nelson Goodman (1966). This is the view that the only things that exist are individuals, in some fairly restrictive sense intended to capture *concrete* things as opposed to *abstract* entities like sets. Basic to Goodman's conception of an individual is the notion of a part and the idea of combining parts to yield wholes. This might smack rather of set theory with its hierarchical collections of objects, but in fact is far more restrictive. Consider for instance three objects, which we will suppose indivisible for the moment: a CUP, a SAUCER and a TABLE. For Goodman there are seven individuals here: they comprise the three objects taken singly, the three pairs such as CUP+SAUCER, and the triple CUP+SAUCER+TABLE. Set theory on the other hand tells us that there are a huge infinity of individual entities present: they include for instance the sets {{}, {CUP, SAUCER}, CUP, {TABLE}} and {{{CUP}}}. Indeed, set theory has no need at all of concrete individuals to start on its transfinite hierarchy of sets; for we can imitate the integers starting from the null set: {}, {{}} etc., and the whole of mathematics follows.

Goodman's wholes – sums of parts – have none of the internal hierarchical structure which set theory insists must be present in every collection, and of

* To avoid the paradoxes, this needs to be restricted in some way. One takes one's pick from the many possibilities. There are always a sufficiently large number of sets to cause trouble however.

which its notation provides a ready analog. To insist on the presence of this irrelevant structure seems futile. Thus in this light Goodman's 'calculus of individuals' seems preferable to set theory. But it also has its faults. *Everything* is an individual, for Goodman: this is the nominalists' basic ontological commitment. Whatsoever individuals A and B may be, A+B is another. Even such unlikely combinations as CUP+THE STATE OF UTAH are allowed. (Of course the same objection applies in greater degree to set theory: but let us leave set theory for a while.) For Goodman's purpose this is entirely appropriate: but for the robot it is most surely not. Some constraints are essential. The robot needs to be able to consider a certain whole as an individual when it is more useful to do so than to regard it as a collection of interacting parts: for some combinations this will never be the case.

This idea of an individual is inspired by Simon's definition of a 'complex system' (Simon 1969). In Simon's phrase, the robot needs to be a *pragmatic holist*: when a certain whole – sum of parts – behaves in a way which is more economically described in terms of it as a unit than in terms of its parts and their interactions, then the robot needs to be able to regard it as indeed a unit, a single individual, an unstructured whole. The importance of this idea for the robot's theory is its entailing that what is considered to be an individual may vary depending upon the robot's circumstances and motivation: individuals come apart and recombine in different ways at different times. I know of no ontological system which faces this problem squarely.

The other major fault, for us, in Goodman's system, is its total lack of mathematical expressive power. Since *every* totality is unstructured, and there are *no* abstract objects – no sets, functions, numbers, relations – it is all but impossible to express facts of number, indeed to do even very elementary mathematics. This is intolerable: some minimal level of mathematical expressiveness is essential for everyday reasoning.

This points back to set theory – but not all the way. It seems anticlimactic to suggest a compromise at this point, but compromises are common in engineering. The robot needs some – I think a *very* little – of the expressive power of set theory (or higher-order logic). I have in mind something like the weak second-order theories of linear order (Lauchli 1968, Siefkes 1968), which have considerable expressive power, but are so weak as to admit decision procedures. It also needs some restriction of the pragmatic kind described above on what constitutes an individual. In general, following McCarthy (1968) and Minsky(1968), we can agree that the robot cannot reason about anything unless it can be told it: but, we can add, it must not be told too much.

There are other important and difficult ontological questions which have been widely debated in modern philosophy and which cut across the platonism–nominalism dispute; most notably perhaps the question of tense. Individuals (in the common sense of material objects) are created and destroyed within time. Consideration of past and future existents raises problems which have been grappled with since Aristotle, with but scant success until fairly recently. I will take up this point again later.

The above may have left the reader pessimistic; but one can make some positive suggestions towards a suitable ontological system for a robot.

The chief fault of both platonism and nominalism is that they are reductionist in tendency: they strive to explain the whole world in terms of a small collection of primitive ideas. The stress is on *conceptual* economy (cf. Quine (1960) on 'the ordered pair as philosophical paradigm'): *structural* economy is left by the wayside.

Now the robot needs quite the opposite. As I have argued above, structural economy is of the greatest importance: the space through which the theorem-prover is to search must be kept to a manageable size. Conceptual economy however is not at all important: on the contrary, conceptual *richness* is, within fairly precise limits, a desirable property of the robot's ontology. Briefly, its individuals should be classified into different kinds, should in fact be *sorted.*

Not only does this simplify the job of its sensors, it helps to reduce the size of the search space still further. Placing a sort structure on the theory amounts to giving a semantically based classification of the syntax, which can be used to control the search. It makes available more information to the syntactic heuristics, for instance. This observation is again familiar to those working in theorem-proving (Guard *et al.* 1969).

This does not contradict what was said earlier about the need for economy in ontological commitment. It is a 'horizontal' rather than a 'vertical' richness which is wanted; many different kinds of individual; but not many of each kind. This is precisely the opposite type of economy from that which reductionist ontologies offer.

There are other reasons for finding a sort structure desirable. Somehow the phenomena which its sensors detect must lead the robot to believe facts about its environment; and it seems likely on the face of it that this mysterious process will inevitably give rise to conceptual classifications of the type being discussed. Thus, perhaps an *event* is distinguished from a *thing* by being intangible and of short duration; a *sound* is reported by a different sense organ than a *colour.* Both physical and phenomenological entities are automatically rendered into sorted language by reason of their different relationships to the sensory input. (This also takes care of the Dreyfusian objection that by inventing and supplying to the robot a sort structure we are somehow cheating, by building in our own insights. The sorts arise from the engineering requirements, not from a synthetic *a priori.*)

The interesting problem is now open to elucidate the sort structure which the robot is to use. Some initial candidates for useful classifications come immediately to mind: *things* (in the ordinary sense of material objects with a comparatively long life span), *people* [*things* capable of executing strategies. The distinction is important because *people* are potential antagonists, and one may need to use game theory against them. *See* (Braithwaite 1955)], *actions* (Davidson 1967) argues for their inclusion. Consider 'he flew to the moon/quickly/in a rocket/last Tuesday/...' Qualifications may be added endlessly. The only way to make logical sense of this is to introduce an entity which has all these properties: thus the above becomes something like $\exists x(flying(x) \wedge subject(x) = him \wedge destination(x) = moon \wedge quick(x) \wedge vehicle(x) = rocket \wedge date(x) = last\ Tuesday \wedge \ldots)$. What is x here but an action – 'his flying to the moon'? Also consider the statement 'he did it quickly'. Did what? – the answer must be, an action.) *events* (again argument due to Davidson (1967). Statements of causality – 'the stabbing caused Caeser's death' – are naturally represented as statements of relationships between events.) *strategies* [a generalizations of *actions*; of especial use in dealing with unknown quantities and with *people*, *see* above and McCarthy and Hayes (1969)]. Other possibilities include *time-instants* and *places*, and of course one will think of others.

The collection of these sorts will have a certain structure: some sorts are included in others, for instance. If the sorts are disjoint, then present-day theorem-provers will deal with them without modification: if they are partially ordered by inclusion, the necessary modifications are not difficult (Guard *et al.* 1969). If however more complicated sort structures are envisaged [as for instance have been devised the better to model natural languages (Bar-Hillel 1950)], then the problem is open, and appears quite difficult.

One other sort which has been used is that of a *possible world* or *situation* (McCarthy 1968, Rescher and Garland 1968, Lewis 1968). The resulting class of theories is very large and contains theories mimicking in their structure almost all the known modal logics. These latter seem to be extremely useful in constructing real-world theories containing statements about tense, knowledge, etc. (see below for instance), and the situation calculus is therefore similarly useful. It might be thought that the ontological commitment involved here was far greater even than in set theory – a 'possible world' is apparently a far more intractable object than a set, and the methods of constructing new worlds from old correspondingly difficult to describe. But the situation calculus is in fact quite tractable (Green 1969), on present form. The reason is that the ontological commitment is not what it seems, since in present-day applications of the situation calculus very little structure is imposed on situations. There are perhaps a few partial orderings defined by means of functions representing actions. Thus the *actual* commitment is only to certain entities having this minimal amount of structure. If one were to construct, within the situation calculus, an axiomatic theory in which many complex interactions between situations were described, then the commitment would be to more complex entities and the resulting theory would indeed be intractable in precisely the sense described above. Attempts to construct more complex theories within the situation calculus have run into just this problem.

One can illustrate this by the process of translating modal logics into the situation calculus. The basic idea is simple and well known. Propositions map into fluents with a single free situation variable: $p \rightarrow p(s)$. The translation is direct ($p \wedge q \rightarrow p(s) \wedge q(s)$ etc.) until one reaches the modal operators. The statement of necessity, $\Box p$, maps onto $\forall t(R(s, t) \supset p(t))$ where R is a binary relation ('alternativeness') between situations. The axioms of the modal logics translate into conditions on R.

So far all is fairly simple. But suppose we try to translate a modal *predicate* calculus into the situation calculus. Then the quantifiers translate as follows:

$$\forall x P(x) \rightarrow \forall x(\text{IN}(x)(s) \supset P'(x)(s))$$

where P' is the translation of P and IN is a new monadic fluent (intuitive meaning: x exists in situation s) for which many axioms must be supplied.

Of course analogous problems arise in considering the modal logics directly, but hidden in the very complicated semantics. The translation into first-order logic via the situation calculus has the merit of bringing them out into the open, so to speak; but it does not eliminate them. There has been lively debate among philosophical logicians as to which approach to the (very real) problems of referential opacity is most rewarding: *see*, for instance, Quine (1953) and more recently Massey (1969) for some broadsides against modal logic.

We will take up this matter again later.

2. THE CHOICE OF LANGUAGE

A theory must be phrased in some formal language for which the robot has available an efficient proof procedure. It is natural to ask: which such language is the most appropriate?

Up to now the most successful robot has used the Resolution formulation of classical first-order logic (Green 1969). One may therefore question the need to look beyond first-order logic: and one would be supported in this scepticism by some eminent philosophers, notably Quine.

However it does seem that there are some good reasons for so doing even though we have no mechanizable inference systems available yet for other languages. In this section I will try to outline three of the most pressing of these reasons. They are concerned with problems respectively of *time*, *perception*, and *ambiguity*.

2.1 Time and tense

The robot, like ourselves, lives in time: it has a future and a past. Like ourselves, its memory will be imperfect and its predictive power even less perfect. Like ourselves, it must have a coherent way of reasoning about the contingent future. This is extremely difficult in classical truth-functional two-valued logic.

The difficulty is not simply the problem of constructing a logic of tenses or time-intervals. The construction of such logics has been a major task in philosophical logic and a huge amount of work has been done (*see*, for instance, Prior 1968). Tenses turn out to be modal operators because they are not truthfunctional (*p*s being true in the future in no way depends upon its being true in the past). It is therefore possible to translate them into the situation calculus roughly as described above, thus rendering everything back once more into classical first-order logic. This has, as noted above, been urged by several authors. However it seems to me that it would be a mistake to conclude that modal logics were useless. The modal logic formulation seems more likely to admit an efficient proof procedure. The reason is that when the modal formulae are translated into the situation calculus, formulae may be inferred from them which are not the translation of any statement of the modal logic. The mapping of search spaces induced by the translation is an imbedding, not a homomorphism.

The difference between the two approaches can also be understood as the difference between using *tenses*, or using quantification over *time-instants*, to handle time. The former are definable in terms of the latter, as above: and in fact the reverse is also true, provided that a clock is available (Prior 1968).

My feeling is that the economy achieved by using tensed language more than compensates for its lack of expressive power, especially in view of this fact that the power can be regained by an artificiality. For what it is worth, one can point to the fact that people are built this way: we seem to have a direct primitive intuition of pastness and presentness (tense), and can only reconstruct time instants by using clocks. Natural language also uses tenses rather than time-instants as its basic way of dealing with time.

There is a deeper difficulty in handling time, however. Any assertion in the future tense is a prediction, and may turn out, in view of the robot's limited predictive ability, to be mistaken. Moreover some statements in the future tense are open – the robot has no opinion because it is unable to find sufficient evidence either for or against them. (This phenomenon occurs of course without introducing tenses, since the robot has limited deductive power in general. But, as argued above, it should in general happen infrequently: otherwise the robot is simply inadequate to its assigned tasks. With regard to *time*, however, the robot cannot help but have an inadequate theory in this sense. The interactions of everyday environments are too complex for any fast predicting mechanism to be adequate.) All these considerations point to the need for a many-valued logic. A minimum of three values – true, false, indeterminate – seems a necessity. There is no efficient way of translating many-valued logics into two-valued logic.

This is argued (with reference to people, not robots) in greater depth by Lukasiewicz (1967). Indeed it was his considerations of future contingency which led to the first construction of a many-valued logic.

The logic of three truthvalues has been put into a mechanizable form [which could however bear some improvement (Hayes 1969)], so we are a little way forward on this front.

However, the major problem here may be an extra-logical one, namely, describing precisely how the truth values change as the robot moves through time. This is closely related (thinking takes time) to what has come to be called the frame problem (McCarthy and Hayes 1969, Minsky 1968) and to the philosophical problems of counterfactual conditionals (*see* McCarthy and Hayes 1969). It remains a central difficulty, but a subproblem which has been partially solved is that of determining which statements are valid in such an environment, i.e., which statements need never be re-examined to see if they have changed their truthvalue.

It might be thought that every valid formula of first-order logic was such a law: but if one wants the set of all such statements to be closed under deduction (so that they form an 'inner logic'), and if one insists – as is extremely natural – that the ordinary connectives are truth-functional, then this is no longer the case.

I will go into this in more detail below: for the present it seems clear that the problems of dealing with time take one beyond first-order logic, and to establish this was my intention.

2.2 Perception

Facts about the robot's environment must be obtained through its sensors. This perceptual process is probably the least understood of all, at present. However one thing stands out in all work which has been done on robot construction, that being the need for conceptual feedback during the perception process.

This need has been noted by workers at Stanford (Feldman *et al.* 1969) and Edinburgh (Murphy 1969). It is also well known in experimental psychology (Averbach and Coriell 1961, Denes 1967, Yarbus 1967) that human perception is not a purely passive process, but is guided by attention mechanisms which select, on the basis of concepts formed from a partial analysis of the input, which parts to attend next.

I will not attempt to describe this process in any detail. Its effect on the language is however profound. Sensors do not deliver perfect information, but rather guesses together with an estimate of their reliability. Implementing the feedback process mentioned above requires that these guesses, and the estimations as to their truth, must be capable of being represented in the language. Moreover the robot must be able to perform inferences on the basis of these facts, inferences which direct the attention of the sensors to the vital parts of the input. These deductions must not be such as to propagate errors in an uncontrolled way: the proof procedure must deal properly with the guesses and the reliability estimates. This implies using a many-valued logic.* Thus we are again forced away from classical logic.

* The SRI group have toyed with this idea, apparently for a similar reason (*see* Green 1969).

It may be objected that the feedback process should be thought of in isolation from the deductive theory, as a sort of complex preprocessing. And in fact this is what has been suggested to date. But there are two reasons why this seems wrong. One is methodological: the division is artificial, and I venture to predict that the generality and power of theorem-proving methods will greatly advance cognitive research when they are applied to it, as has happened to other fields in the past. The other is more important. In a fast-moving (compared to its operating speed) environment the robot will need to *act* on the basis of perceptual half-truths, will need, for instance, to have protective reflexes when it detects severe threats, even though the threat may be only partly verified by the sensors. This requires that the whole deductive apparatus be intimately associated with the sensory feedback at every stage.

There are other problems which arise when one deals with percepts, however, by whatever process they are produced.

Consider the assertion $P(a)$: it says that the individual a enjoys the property P. What does it mean to assign a truthvalue of (say) 0·5 to this statement? It could mean that the individual a has the property P to this degree – that a is '*half-P*'; or it could mean that P is *definitely* true of an entity which is '*half-a*'; which has been but partially identified and can only be called a with limited confidence. The former interpretation is the usual one in many-valued logics, but the latter – which is far more damaging to the formal semantics and the conceptual basis than the other – seems to be what is wanted here.

Similar difficulties arise in interpreting modal assertions (is $\Diamond P(a)$ an assignment of *possibly-P* to a, or an assignment of P to a '*possible-a*', whatever that means?) and modal logicians have been active in investigating them. It seems that the only way to construct a coherent semantics for modal predicate logic is to delineate precisely what the criteria for identifying individuals are. It is impossible to summarize the whole of this work here, but it is clearly directly relevant to the perceptual difficulties being discussed.

The distinction between individual constants and variables, relatively minor in classical logic, is crucial in modal logic. Variables, being simply cross-referencing devices for quantification, are not susceptible to these interpretive difficulties. But constants are encumbered by their referential apparatus and their freedom of use is thereby restricted. Thus such classically valid inferences as $\forall x P(x) \rightarrow P(a)$ typically fail in quantified modal logics.

These sort of complications have led some authors, notably Quine, to attack quantified modal logic as requiring unsavoury philosophical assumptions; but for the present purpose this is precisely the sort of behaviour one would want, and this research into the philosophical foundations of modal logic seems certain to produce valuable results for the robot's cognitive difficulties.

2.3 Ambiguity

In natural language usage one continually uses nouns which mean slightly different things depending on the context of their use. To take a singularly trivial example, 'pass the salt' usually is a request to pass a container of salt rather than the salt itself: the context is sufficient to force the reinterpretation of the word 'salt'.

That this sort of ambiguity does not cause problems is due to the fact that the recipient of such a message is himself capable of disambiguating it: the receiver is an intelligent creature.

Disregarding the question of communication for now, it seems likely that the robot's planning facilities will comprise a hierarchy of some sort. There will be higher-level plans defining strategies which are interpreted by lower-level planning devices whose responsibility it is to deal with the tactics of the situation. Now it seems reasonable that the sort of ambiguity being discussed should be present in the plans formed by (and hence, in the cogitations of) the higher-level control, since these plans are to be interpreted by an 'intelligent' device of some sort.

I have not yet, however, said why such ambiguity would be desirable. It is, as usual, a matter of efficiency. Suppose one undertook never to commit the unFregean sin of using one name to refer to different objects in the sort of way indicated. Then his theory must contain many distinct names for closely related entities, and in order that their close relationship be apparent, it must also contain axioms concerning these interrelationships. Thus his theory will be cluttered up with what might be called the micro-physics of the situation. It is *essential* that this kind of close analysis be left to the tactical – lower-level – planning stage.

Now, the presence of this ambiguity strikes at the very root of classical logic. It is closely related to the modal difficulties described above: in both cases we have a crucial weakening of the *denotation* relationship holding between the objects of reference (individuals) and their names (individual constant symbols). What seems to be necessary is that some flexibility is available in ontological commitment: in deciding what comprises an individual. Again we are brought back to the nominalistic difficulties described earlier.

2.4 Modal versus classical

The two kinds of nonclassical logic which arise naturally are seen to be modal logics and many-valued logics. As mentioned briefly above, it is possible to translate both of these back into classical first-order logic in various ways. It is therefore not clear whether one has to extend classical logic in constructing robot-oriented theories.

In the case of many-valued logics the mapping is clearly unattractive. One introduces constants $t_1, t_2 \ldots$ which are supposed to denote the various truthvalues; and then instead of asserting a proposition p with truth-value

t_n, one asserts the classical (two-valued) truth of '$p = t_n$'. This essentially amounts to constructing a first-order axiomatic metatheory of the many-valued logic. Although an interesting theoretical exercise, the resulting theory is less intuitive, far clumsier, and is likely to be hopelessly less efficient than the original logic.

In the case of modal logic the matter is less clear-cut, however. I have touched upon this famous dispute several times already. The full story seems to be as follows. The situation-calculus modelling of modal logic is sometimes intuitively clearer than the modal logic itself (tense logic), sometimes less so (epistemic logic). It is likely that efficient proof procedures for languages involving modal notions will have to treat modal operators – or the first-order constructions which mimic them – in a special way. This seems, in view of the imbedding of the search spaces mentioned above, easier to do in the context of the modal logic itself. The situation-calculus approach, on the other hand, has the methodological advantage of providing a unified framework within which to relate different modal notions, but it does not eliminate the real difficulties of elucidating the referential structures implicit in the semantics. Indeed the construction of a suitable first-order theory and the construction of a formal semantics for the modal calculus seem isomorphic problems (see Montague (1968) for an example of the latter).

It seems clear that each approach has its advantages and snags. The only real sin is dogmatism.

There are other reasons for considering nonclassical logics, especially the problems of reasoning about *knowledge*, and connexions between the robot's reasonings and his actions. I have not discussed the former since it was aired at some length in an earlier paper (McCarthy and Hayes 1969) and I have nothing new on the subject; and I do not understand the latter, but expect that problems analogous to those of perception will arise.

Of the three reasons outlined, the first – the difficulty with time and tense – seems the most fundamental. To some extent it subsumes the frame problem and the perceptual difficulties, since both of these arise in the context of a discovery process which takes place in time. It is also the area where most progress has been made, due probably to its philosophical interest. In the next section I will outline some recent work in philosophical logic which is relevant to it.

ASSERTIONS IN TIME

The results described in this section are due to Kripke (1963, 1965) and Grzegorczyk (1964, 1968). Also much of the underlying philosophical discussion is taken from these papers. There will be complete confusion between use and mention.

The robot's life can be thought of as a process of discovery. It finds out new facts as time passes. It is capable of perceiving, and choosing from, alternative courses of action. Moreover when it has acted upon such a choice it cannot retrace its steps. Disregarding for the moment the purpose of such actions, we will isolate those which result in the acquisition of new data about the environment.

The natural formal representation of this process is as a directed graph, or *digraph*. At the nodes of the graph will be collections of statements representing the state of the robot's knowledge at successive time-instants. The arcs of the graph correspond to experiments which yield new information about the environment. The directedness of the graph reflects the irreversibility of the robot's decisions.

Let $\leqslant$ denote the ordering relation on such a digraph, and let N be a finite or infinite set of nodes. At each node $n \in N$, there are sets I_n of *individuals* and A_n of ground atomic statements (*atoms*). I_n is a set of individuals which the robot has met at that time, and A_n is a set of observational facts which the robot has verified.

In order to make this quite precise we must specify the language which the robot is to use to represent this knowledge. This will be, for the moment, the classical first-order predicate calculus augmented by the modal operator G. The intended meaning of G is '*it will always henceforth be true that*...'. We may define the dual modal operator F as $\neg G \neg$, and then F means '*it will be true that*...', i.e., F is the future tense operator. These modal operators comprise the robot's language for speaking of time.

The digraphs, with their associated structure, act as interpretations of this language in the way described in detail below and for this interpretation Kripke gives a completeness result:

Theorem I.

The set of formulas true at all nodes of all digraphs is the set of theorems of the modal logic CS4. (See appendix.)

The interpretation mapping is defined as follows:

The truthvalue of an atom A at a node n is T if $A \in A_n$; otherwise the truthvalue of A at n is F. Notice that we assign F to A even though neither of A, $\neg A$ is true at n. This asymmetry between truth and falsity will be important in what follows. We can now define the truthvalues of more complicated formulae by induction on their structure. Let $[P]_n$ be the truthvalue of P at n. Then:

C1. $[\neg R]_n = \neg [R]_n$

C2. $[P \vee Q]_n = [P]_n \vee [Q]_n$

C3. $[\exists x P]_n = T$ if $[P(a/x)]_n = T$ for some $a \in I_n$, otherwise $= F$.

C4. $[GA]_n = T$ if $[A]_m = T$ for every $m \geqslant n$, otherwise $= F$.

The first three conditions mean that the interpretation of formulae not containing G is the usual classical one, based on the evidence available. Condition C4 captures the intended meaning of the modal operator G: GA is true at n if and only if A is true throughout n's future (which includes n itself).

It seems to me that this result is of central importance for robot research. The digraph semantics generalizes the common notion of 'look-ahead tree' of classical economic theory and much AI research, but the above result does not in any way depend on the digraph's structure being accessible to the robot. This assumption of perfect information is the main weakness of the use of the 'look-ahead' paradigm in these fields (Simon 1967).

This result can be extended in several ways. The construction of the digraphs has so far been perfectly free: no conditions have been placed upon the sets I_n and A_n. Thus individuals may be created and destroyed, or may change their properties in arbitrary ways, as time progresses. This seems appropriate for everyday reasoning in a changing world. However, one may ask for a similar construction which would be appropriate for the discovery of laws, that is facts, about the world which are not expected to change.

To achieve this we impose the following restrictions on the digraph construction:

R1: if $n \leqslant m$ then $I_n \subseteq I_m$

R2: if $n \leqslant m$ then $A_n \subseteq A_m$.

These say that the universe is growing, that is, that individuals once discovered never disappear; and that atomic facts never change. These seem to be the appropriate conditions for law discovery, but one can make out a case for their being true in general. This requires the reinterpretation of the atoms and it will be discussed more fully later.

A digraph obeying R1 and R2 will be called a *law-graph*. Using the same definition of the truthvalue of an atomic formula as above, we need a suitable collection of induction clauses to define the truthvalues of more complicated formulae. Bearing in mind the asymmetry between truth and falsity mentioned earlier, it is easily seen that the clause for negation must be strengthened in order to be sure that negative statements shall also be law-like. Also, since we want *modus ponens* and instantiation to be correct rules of inference in the resulting logic, we must pay particular attention to the clauses for implication and the universal quantifier. In the latter case, for instance, the assertion of $\forall x P(x)$ must only be allowed when it is guaranteed that it cannot be falsified in the future – even though new individuals are discovered – for otherwise it cannot be regarded as a law.

The resulting collection of induction clauses is then the following:

L1. $[\neg P]_n = T$ if $[P]_m = F$ for every $m \geqslant n$, otherwise $= F$.

L2. $[P \vee Q]_n = [P]_n \vee [Q]_n$.

L3. $[P \wedge Q]_n = [P]_n \wedge [Q]_n$.

L4. $[P \supset Q]_n = T$ if $[P]_m \supset [Q]_m$ for every $m \geqslant n$, otherwise $= F$.

L5. $[\exists x P]_n = T$ if $[P(a/x)]_n = T$ for some $a \in In$.

L6. $[\forall x P]_n = T$ if $[P(a/x)]_m = T$ for every $m \geqslant n$ and every $a \in A_m$, otherwise $= F$.

L7. $[GA]_n = T$ if $[A]_m = T$ for every $m \geqslant n$, otherwise F.

(Notice that L2 = C2, L5 = C3 and L7 = C4.)

With this semantics Kripke (1965) proves the following result (*see* also Grzegoroczyk 1964, 1968):

Theorem 2

The set of statements (strictly: the set of statements not containing G) true at all nodes of all law-graphs is the set of theorems of the intuitionistic predicate calculus IPC.

As remarked earlier, it is perhaps surprising that one gets a weaker logic than the classical predicate calculus. Consideration of a simple counter example to the law of excluded middle however shows why this must be the case. The simple law tree in question is shown in figure 1. It is easily seen that $P(a)$ is F at n, since it certainly is not T. But, $\neg P(a)$ is F also at n, since it is F at m. Therefore $P(a) \vee \neg P(a)$ must be F at n.

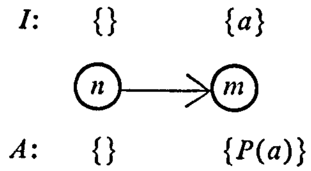

Figure 1

The reason for this behaviour is clearly the insistence upon two-valued truthfunctionality. It may be that one could avoid classifying the law of the excluded middle as a falsehood by introducing a third truthvalue; and indeed this would fall under the general plan outlined earlier. But the law could still not be classed as a truth without sacrificing truthfunctionality. Attempting to accommodate this within a modal logic leads to inconsistency, as follows from results of Montague (1963). I feel that this kind of behaviour should simply be accepted.

Of course, the robot will never find a *counterexample* to the excluded middle, or indeed to any classically valid statement: but that is another question (see Theorem 3 below).

One may quarrel with the condition R1 on law-trees, even in the context of law discovery, on the grounds that *individuals* must be allowed to decay, albeit that *facts* may not change. For where are eternal individuals found? They certainly cannot be physical objects in everyday experience.

There are two answers to this objection. One is that it is appropriate for the robot to regard as eternal any object of sufficient stability that it has a *long* life expectancy: houses, geological features of the landscape, etc. The resulting oversimplification will be true most of the time; and on the few occasions it is not, it is acceptable for the robot to spend time on a lengthy revision of its world-model. It may be indeed that this will work – people seem to have similar problems.

The second answer to the objection is more far-reaching. It is to regard existence as including past existence. This goes part way towards the 'de-tenser' attitude (Massey 1969) that the only meaningful use of 'exists' is *timeless*: thus one has to speak of existence *at a certain time* to capture the meaning of 'exists' in tensed language. The suggested compromise amounts to accepting this dogma for the past, but retaining the tense-logical view for the future. Thus '$\exists x P(x)$' now reads 'there is, *or has been*, an individual x such that $P(x)$'. To keep the expressive power of the old theory one must introduce a one-place predicate *dead*(x): *dead*(a) is to be true at n when $a \in I_n$ but a does not denote any individual which actually exists – in the tensed sense – at n. Then the assertion of *present* existence becomes $\exists x$ $(\neg dead(x) \wedge \ldots)$. In this way one satisfies condition R1, trivially.

In order to render all time-logic into the law-tree semantics one can go further and assume that the atomic propositions are *dated*: that they include a reference to the node n at which they are asserted. This is even more of a step towards the 'de-tenser' view of time, but again only with reference to the past; the future is still to be handled using G. This makes condition R2 trivially satisfied also, but at considerable cost in brevity of the formalism and memory requirements.

This asymmetry between past and future, although probably repugnant to a philosopher, seems very appropriate to the robot. Its memory will be far more reliable than its predicting mechanism; moreover it can choose which facts to forget, but it cannot completely choose which to predict.

Theorems 1 and 2 are closely related. There is in fact a deduction-preserving mapping (*see* Appendix) from IPC into CS4 under which Theorem 2 becomes a corollary of Theorem 1.

Grzegorczyk has extended this result in another direction. Still considering law-trees, one can ask for the logic of hypotheses. Let us say that an assertion is *supposable* at a node n if its negation does not follow from the set of ground facts A_n, together with a fixed consistent set of statements L (a set of underlying laws which are taken as given). One can ask, what statements are always *supposable*?

Grzegorczyk (1968) proves

Theorem 3

If L contains the theorems of IPC, then the set of statements supposable at all nodes of all law-trees contains all theorems of the classical predicate calculus. And

Theorem 4

If L is empty, then the set of statements supposable at all nodes of all law-trees contains all theorems of the system of strict entailment of Anderson and Belnap (1962).

(Theorem 4 is weak, since the set certainly contains other assertions also.)

Theorem 3 makes precise the feeling that classical logic should somehow be 'correct': there will never be a counterexample to any thesis of it. Note however that in general the set of statements supposable at a given node is *inconsistent*: at node n in figure 1, for instance, it contains both $P(a)$ and $\neg P(a)$.

It seems to me that a clear case exists for regarding CS4, or its situation-calculus counterpart, as a good candidate for the basic logic for use by a robot. This is not a restriction upon the use of classical reasoning: CS4 (unlike IPC) contains classical predicate logic as a subsystem. But it also contains a mechanism for handling future contingency which, by the above results, seems exactly appropriate for the robot's situation.

CODA

Throughout this paper I have tried to use results in, and arguments from, philosophical logic and analytical philosophy. It seems to me that these fields have more to offer artificial intelligence than is generally realized. True, the philosopher's motivation in constructing, for instance, a phenomenalistic theory within first-order logic, may be different from that of the AI worker: nevertheless he will have to consider similar problems (consider the quotation from Strawson in the introduction for instance), and to the extent that he does, so will his results be relevant.

Both the analytical philosopher and the designer of intelligent software are doing what might be called 'mental engineering': constructing precise, formal models of some aspects of intelligent thought or behaviour. There is a wide gap between them, but it is narrowing. On the philosopher's side, through the use of the analytical tools of formal logic; on the AI side, through the development and use of efficient theorem-proving programs, and the linguistic approach to problem-solving in general (Minsky 1968). Even methodologically there are similarities between the two fields: compare for instance the use of 'toy universes' by Strawson (1959) and Goodman (1966) to that by Toda (1965) and Doran (1969).

I hope that we will see more co-operation between AI and philosophy in the future. The traffic would not be one-way: for instance, consideration of the perceptual feedback mentioned in section 2.2 above may perhaps throw new light on the longstanding phenomenalist/physicalist controversy.

Acknowledgements

This work was supported by the Science Research Council.

I have already remarked on my indebtedness to John McCarthy and Cordell Green. Some of the other people who have helped me in various ways are Bruce Anderson, Rod Burstall, Jim Doran, Bob Kowalski, Bernard Meltzer, Donald Michie and Alan Robinson.

In particular, many of the ideas in section 1 arose from conversations with Bob Kowalski.

REFERENCES

Allen, J. & Luckham, D. (1970) An interactive theorem-proving program. *Machine Intelligence 5* (eds Meltzer, B. & Michie, D.).

Anderson, A.R. & Belnap Jr., N.D. (1962) The pure calculus of entailment. *J. Symbolic Logic*, **27**, 19–52.

Averbach, E. & Coriell, A.S. (1961) Short term memory vision. *Bell System tech. J.*, **40.**

Bar-Hillel, Y. (1950) On syntactic Categories. *J. Symbolic Logic*, **15**, 1.

Braithwaite, R.B. (1955) *The Theory of Games as a Tool for the Moral Philosopher.* Cambridge University Press.

Davidson, D. (1967) Casual relations. *J. of Philosophy*, **64**, 21.

Davidson, D. (1967) The logical form of action sentences. *The Logic of Decision and Action* (ed. Rescher, N.). University of Pittsburgh Press.

Denes, P.B. (1967) On the motor theory of speech perception. *Models for the perception of speech and visual form* (ed. Wahten-Dunn, W.). Cambridge, Mass.: MIT Press.

Doran, J. (1969) Planning and generalization in an automaton/environment system. *Machine Intelligence 4*, pp. 433–54 (eds Meltzer, B. & Michie, D.). Edinburgh: Edinburgh University Press.

Feldman, J.A., *et al.* (1969) The Stanford Hand-Eye project. *Proc. First Int. Conf. on Artificial Intelligence.* Washington.

Goodman, N. (1966) *The Structure of Appearance.* New York: Bobs-Merrill Co. Inc.

Gould, W.E. (1966) A matching procedure for *w*-order logic. *Scientific Report No. 4.* Princeton: Applied Logic Corporation.

Green, C. (1969) Application of theorem-proving to problem-solving. *Proc. First Int. Conf. on Artificial Intelligence.* Washington.

Grzegorczyk, A. (1964) A philosophically plausible formal interpretation of intuitionistic logic. *Indagationes Math.*, **26**, 596–601.

Grzegorczyk, A. (1968) Assertions depending upon time and corresponding logical calculi. *Compositio Mathematica*, **20**, 83–7.

Guard, J.R. *et al.* (1969) Semiautomated Mathematics. *J. Ass. comput. Mach.*, **16**, 1.

Hayes, P. (1969) A machine-oriented formulation of the extended functional calculus. *A.I. Memo*, A.I. Project. Stanford University.

Kleene, S.C. (1952) *Introduction to Metamathematics.* Princeton, New Jersey: Van Nostrand.

Kripke, S. (1963) Semantical analysis of modal logic *I. Zeitshr. f. math. Logik und Grundlagen d. Math.*, **9**, 67–96.

Kripke, S. (1965) Semantical analysis of intuitionistic logic I. *Formal Systems and Recursive Functions* (eds Crossley, N. & Dummett, M.). Amsterdam: North Holland.

Lauchli, H. (1968) A decision procedure for the weak second-order theory of linear order. *Contributions to mathematical logic* (eds Scmidt, H.A., Schutte, K. & Thiele, H-J.). Amsterdam: North Holland.

Lewis, D.K. (1968) Counterpart theory and quantified modal logic. *J. of Philosophy*, **45**, 5.

Lukasiewicz, J. (1967) On determinism. *Polish Logic* 1920–1939 (ed. McCall, S.). Oxford: Clarendon Press.

Massey, G.J. (1969) Tense logic! Why bother? *Nous*, **3**, 17–32.

McCarthy, J. (1964) A tough nut for proof procedures. *A.I. Memo*, A.I. Project, Stanford University.

McCarthy, J. (1968) Programs with common sense. *Semantic Information Processing* (ed. Minsky, M.L.). Cambridge, Mass.: MIT Press.

McCarthy, J. & Hayes, P.J. (1969) Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence 4*, pp. 463–502 (eds Meltzer, B. & Michie, D.). Edinburgh: Edinburgh University Press.

Minsky, M.L. (1968) Descriptive languages and problem solving. *Semantic Information Processing* (ed. Minsky, M.L.). Cambridge, Mass: MIT Press.

Montague, R. (1963) Syntactical considerations on modal logic. *Acta Phil. Fennica*, **16.**

Montague, R. (1968) Pragmatics. *Contemporary Philosophy* (ed. Klibansky, R.). Firenze: La Nuova Italia Editrice.
Murphy, A.S. (1969) An application of heuristic search procedures to picture interpretation. *Memorandum MIP–R–61*, Dept. of Machine Intelligence, Edinburgh University.
Prawitz, D. (1960) An improved proof procedure. *Theoria*, **26**, 102–39.
Prawitz, D. & Malmnäs, P.-E. (1968) A survey of some connections between classical, intuitionistic and minimal logic. *Contributions to Mathematical Logic* (eds Schmidt, H.A., Schutte, K. & Thiele, H-J.). Amsterdam: North Holland.
Prior, A.N. (1968) *Past, Present and Future*. Oxford: Clarendon Press.
Quine, W.V.O. (1953) Mr. Strawson on logical theory. *Mind*, **42**, 435–51.
Quine, W.V.O. (1960) *Word and Object*. Cambridge, Mass.: MIT Press.
Quinlan, J.R. & Hunt, E.B. (1968) A formal deductive problem-solving system. *J. Assoc. comput. Mach.*, **15**, 4.
Rescher, N. & Garland, J. (1968) Topological Logic. *J. symbolic Logic*, **33**, 4.
Robinson, J.A. (1965a) A machine-oriented logic based on the resolution principle. *J. Ass. comput Mach.*, **12**, 23–41.
Robinson, J.A. (1965b) Automatic deduction with hyper-resolution. *Int. J. comput. Math.*, **1**, 227–34.
Siefkes, D. (1968) Recursion theory and the theorem of Ramsay in one-place second-order successor arithmetic. *Contributions to mathematical logic* (eds Schmidt, H.A., Schutte, K. & Thiele, H-J.). Amsterdam: North Holland.
Simon, H.A. (1967) The logic of heuristic decision making. *The Logic of Decision and Action* (ed. Rescher, N.). University of Pittsburgh Press.
Simon, H.A. (1969) *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press.
Strawson, P.F. (1959) *Individuals*. London: Methuen & Co.
Toda, M. (1963) Utilities, induced utilities, and small worlds. *Behav. Science*, **10**, 238–54.
Wos, L.T. Carson, D.F. & Robinson, G.A. (1965) Efficiency and completeness of the set of support strategy in theorem proving. *J. Ass. comput. Mach*, **12**, 536–41.
Yarbus, A.L. (1967) *Eye movements and Vision*. New York: Plenum Press.

APPENDIX

The modal predicate calculus CS4 is obtained by adding to the axioms of the classical first-order predicate calculus the following:

S4 A1 $Gp \supset p$
S4 A2 $G(p \supset q) \supset .\ Gp \supset Gq$
S4 A3 $Gp \supset GGp$

and the additional rule of inference:

S4 R1 If $\vdash p$, then $\vdash Gp$,

where p, q are arbitrary propositions.

The intuitionistic predicate calculus IPC uses the vocabulary of classical predicate calculus, but the quantifiers are not interdefinable and the implication connective is not definable in terms of disjunction and negation.

The axioms and rules of inference are listed in Kleene's monograph (1952) at the top of page 82. IPC is obtained from postulates 1a, 1b, 2, 3, 4a, 4b, 5a, 5b, 6, 7, 9, 10, 11 and 12, together with 8^{I}: $\neg A \supset (A \supset B)$ which replaces 8°. Kleene also gives a Gentzen-type formulation of IPC on page 481.

The mapping from IPC into CS4 is defined as follows (Prawitz and Malmnäs 1968). A formula P maps into P° where:

$$P^\circ = GP \text{ if } P \text{ is an atom}$$
$$(P \wedge Q)^\circ = P^\circ \wedge Q^\circ$$
$$(P \vee Q)^\circ = P^\circ \vee Q^\circ$$
$$(\neg P)^\circ = G\neg P^\circ$$
$$(P \supset Q)^\circ = G(P^\circ \supset Q^\circ)$$
$$(\exists x P)^\circ = \exists x P^\circ$$
$$(\forall x P)^\circ = G\forall x P^\circ.$$

The reader will see that the introductions of G for universal quantifiers, negation and implication correspond precisely to clauses L6, L1 and L4 in the law-graph semantics. The introduction of G for atoms corresponds to the conditions R1 and R2 on law-graphs, which effectively eliminate the distinction between A and GA when A is an atom.