P. J. HAYES

The Logic of Frames

Introduction: Representation and Meaning

Minsky introduced the terminology of 'frames' to unify and denote a loose collection of related ideas on knowledge representation: a collection which, since the publication of his paper (Minsky, 1975) has become even looser. It is not at all clear now what frames are, or were ever intended to be.

I will assume, below, that frames were put forward as a (set of ideas for the design of a) formal language for expressing knowledge, to be considered as an alternative to, for example, semantic networks or predicate calculus. At least one group have explicitly designed such a language, KRL (Bobrow/Winograd, 1977a, 1977b), based on the frames idea. But it is important to distinguish this from two other possible interpretations of what Minsky was urging, which one might call the metaphysical and the heuristic (following the terminology of (McCarthy/Hayes, 1968)).

The "metaphysical" interpretation is, that to use frames is to make a certain kind of assumption about what entities shall be assumed to exist in the world being described. That is, to use frames is to assume that a certain *kind* of knowledge is to be represented by them. Minsky seems to be making a point like this when he urges the idea that visual perception may be facilitated by the storage of explicit 2-dimensional view prototypes and explicit rotational transformations between them. Again, the now considerable literature on the use of 'scripts' or similar frame-like structures in text understanding systems (Charniak, 1977; Lehnert, 1977; Schank, 1975) seems to be based on the view that what might be called "programmatic" knowledge of stereotypical situations like shopping-in-a-supermarket or going-somewhere-on-a-bus is necessary in order to understand English texts about these situations. Whatever the merits of this view (its proponents seem to regard it as simply *obvious*, but see (Feldman, 1975) and (Wilks, 1976) for some contrary arguments), it is clearly a thesis about what sort of things a program needs to know, rather than about *how* those things should or can be *represented*. One could describe the sequence of events in a typical supermarket visit as well in almost any reasonable expressive formal language.

The "heuristic", or as I would prefer now to say, "implementation", interpretation is, that frames are a computational device for organising stored representations in computer memory, and perhaps also, for organising the processes of retrieval and inference which manipulate these stored representations. Minsky seems to be making a point like this when he refers to the computational ease with which one can switch from one frame to another in a frame-system by following pointers. And many other authors have referred with evident approval to the way in which frames, so considered, facilitate certain retrieval operations. (There has been less emphasis on undesirable computational features of frame-like hierarchical organisations of memory.) Again, however, none of this discussion engages representational issues. A given representational language can be implemented in all manner of ways: predicate calculus assertions may be implemented as lists, as character sequences, as trees, as networks, as patterns in an associative memory, etc: all giving different computational properties but all encoding the same representational language. Indeed, one might almost characterise the art of programming as being able to deploy this variety of computational techniques to achieve implementations with various computational properties. Similarly, any one of these computational techniques can be used to implement many essentially different representational languages. Thus, circuit diagrams, perspective line drawings, and predicate calculus assertions, three entirely distinct formal languages (c.f. Hayes, 1975), can be all implemented in terms of list structures. Were it not so, every application of computers would require the development of a new specialised programming language.

Much discussion in the literature seems to ignore or confuse these distinctions. They are vital if we are to have any useful taxonomy, let alone theory, of representational languages. For example, if we confuse representation with implementation then LISP would seem a universal representational language, which stops all discussion before we can even begin.

One can characterise a representational language as one which has (or can be given) a semantic theory, by which I mean an account (more or less formal, more or less precise — this is not the place to argue for a formal model theory, but see Hayes, 1977) of how expressions of the language relate to the individuals or relationships or actions or configurations, etc., comprising the world, or worlds about which the language claims to express knowledge. (Such an account may — in fact must — entail making some metaphysical assumptions, but these will usually be of a very general and minimal kind (for example, that the world consists of individual entities and relationships of one kind or another which hold between them: this is the ontological committment needed to understand predicate logic)). Such a semantic theory defines the *meanings* of expressions of the language. That's what makes a formal language into a representational language: its expressions carry meaning. The semantic theory should explain the way in which they do this carrying. To sum up, then, although frames are sometimes understood at the metaphysical level, and sometimes at the computational level, I will discuss them as a representational proposal: a proposal for a language for the representation of knowledge, to be compared with other such representational languages: a language with a meaning.

What Do Frames Mean?

A frame is a data structure — we had better say *expression* — intended to represent a 'stereotypical situation'. It contains named '*slots*', which can be filled with other expressions — *fillers* — which may themselves be frames, or presumably simple names or identifiers (which may themselves be somehow associated with other frames, but not by a slot-filler relationship: otherwise the trees formed by filling slots with frames recursively, would always be infinitely deep). For example, we might have a frame representing a typical house, with slots called *kitchen*, *bathroom*, *bedrooms*, *lavatory*, *room-with-TV-in-it*, *owner*, *address*, etc.. A particular house is then to be represented by an *instance* of this *house* frame, obtained by filling in the slots with specifications of the corresponding parts of the particular house, so that, for example, the *kitchen* slot may be filled by an instance of the frame *contemporary-kitchen* which has slots *cooker*, *floorcovering*, *sink*, *cleanliness*, etc., which may contain in turn respectively an instance of the *split-level* frame, the identifier *vinyl*, an instance of the *double-drainer* frame, and the identifier '13' (for "very clean"), say. Not all slots in an instance need be filled, so that we can express doubt (e.g. "I dont't know where the lavatory is"), and in real 'frame' languages other refinments are included, e.g. descriptors such as "which-is-red" as slot fillers, etc. We will come to these later. From examples such as these (c.f. also Minsky's birthday-party example in Minsky, 1975), it seems fairly clear what frames mean. A frame instance denotes an individual, and each slot denotes a relationship which may hold between that individual and some other. Thus, if an instance (call it G00097) of the *house* frame has its slot called *kitchen* filled with a frame instance called, say G00082, then this means that the relationship *kitchen* (or, better, *is kitchen of*) holds between G00097 and G00082. We could express this same assertion (for it is an assertion) in predicate calculus by writing: is kitchen of (G00097, G00082).

Looked at this way,frames are essentially bundles of properties. *House* could be paraphrased as something like $\lambda x.$ (kitchen (x, y_1) & bathroom (x, y_2) & ...) where the free variables y_i correspond to the slots. Instantiating *House* to yield a particular house called *Dunroamin* (say), corresponds to applying the λ-expression to the identifier *Dunroamin* to get kitchen (dunroamin, y_1) & bathroom (dunroamin, y_2) & ... which, once the "slots" are filled, is an assertion about Dunroamin.

Thus far, then, working only at a very intuitive level, it seems that frames are simply an alternative syntax for expressing relationships between individuals, i.e. for predicate logic. But we should be careful, since although the meanings may appear to be the same, the inferences sanctioned by frames may differ in some crucial way from those sanctioned by logic. In order to get more insight into what frames are supposed to mean we should examine the ways in which it is suggested that they be *used*.

Frame Inference

One inference rule we have already met is *instantiation:* given a frame representing a concept, we can generate an instance of the concept by filling in its slots. But there is another, more subtle, form of inference suggested by Minsky and realised explicitly in some applications of frames. This is the "criteriality" inference. If we find fillers for all the slots of a frame, then this rule enables us to infer that an appropriate instance of the concept does indeed exist. For example, if an entity has a kitchen and a bathroom and an address and ..., etc.; then it must be a house. Possession of these attributes is a sufficient as well as necessary condition for an entity to qualify as a house, criteriality tells us.

An example of the use of this rule is in perceptual reasoning. Suppose for example the concept of a letter is represented as a frame, with slots corresponding to the parts of the letter (strokes and junctions, perhaps), in a program to read handwriting (as was done in the Essex Fortran project (Brady/Wielinga, 1977)). Then the discovery of fillers for all the slots of the 'F' frame means that one has indeed found an 'F' (the picture is considerably more complicated than this, in fact, as all inferences are potentially subject to disconfirmation: but this does not affect the present point.).

Now one can map this understanding of a frame straightforwardly into first-order logic also. A frame representing the concept C, with slot-relationships $R_1, \ldots, R_n$, becomes the assertion

$$\forall x\ (C(x) \equiv \exists y_1, \ldots, y_n.\ R_1(x, y_1)\ \&\ \ldots\ \&\ R_n(x, y_n))$$

or, expressed in clausal form:

$$\forall x\ C(x) \supset R_1(x, f_1(x))$$
$$\&\ \forall x\ C(x) \supset R_2(x, f_2(x))$$
$$\&\ \ \vdots$$
$$\&\ \forall x y_i\ R_1(x, y_1)\ \&\ R_2(x, y_2)\ \&\ \ldots\ \&\ R_n(x, y_n). \supset C(x)$$

The last long clause captures the criteriality assumption exactly. Notice the Skolem functions in the other clauses: they have a direct intuitive reading, e.g. for *kitchen*, the corresponding function is *kitchenof*, which is a function from houses to their kitchens. These functions correspond exactly to the *selectors* which would apply to a frame, considered now as a data structure, to give the values of its fields (the fillers of its slots). All the variables here are universally quantified. If we assume that our logic contains equality, then we could dispense altogether with the slot-relations R_i and express the frame as an assertion using equality. In many ways this is more natural. The above then becomes:

$$C(x) \supset \exists y.\ y = f_1(x)$$
$$\&\ \text{etc.}$$
$$f_1(x) = y_1\ \&\ \ldots\ \&\ f_n(x) = y_n. \supset C(x)$$

(Where the existential quantifiers are supposed to assert that the functions are applicable to the individual in question. This assumes that the function symbols f_i denote partial functions, so that it makes sense to write $\neg\exists y.\ f_1(x)$. *Other notations are possible.)*

We see then that criterial reasoning can easily be expressed in logic. Such expression makes clear, moreover (what is sometimes not clear in frames literature) whether or not criteriality is being assumed. A third form of frames reasoning has been proposed, often called *matching* (Bobrow/Winograd, 1977a). Suppose we have an instance of a concept, and we wish to know whether it can plausibly be regarded as also being an instance of another concept. Can we view John Smith as a dog-owner?, for example, where J.S. is an instance of the Man frame, let us suppose, and Dogowner is another frame. We can rephrase this question: can we find an instance of the dog-owner frame which *matches* J.S.? The sense of *match* here is what concerns us. Notice that this cannot mean a simple syntactic unification, but must rest — if it is possible at all — on some assumptions about the domain about which the frames in question express information.

For example, perhaps Man has a slot called *pet*, so we could say that a sufficient condition for J.S.'s being matchable to Dog-owner is that his *pet* slot is filled with as object known to be canine. Perhaps Dog-owner has slots *dog* and *name:* then we could specify how to build an instance of dog-owner corresponding to J.S.: fill the *name* slot with J.S.'s name (or perhaps with J.S. himself, or some other reference to him) and the *dog* slot with J.S.'s pet. KRL has facilities for just this sort of transference of fillers from slots in one frame to another, so that one can write routines to actually perform the matchings.

Given our expressions of frames as assertions, the sort of reasoning exemplified by this example falls out with very little effort. All we need to do is express the slot-to-slot transference by simple implications, thus: Isdog(x) & petof(x,y). $\supset$ dogof(x,y) (using the first formulation in which slots are relations). Then, given:

name (J.S., "John Smith") (1)
& pet (J.S., Fido) (2)
& Isdog (Fido) (3)

(the first two from the J.S. instance of the 'man' frame, the third from general world-knowledge: or perhaps from Fido's being in fact an instance of the Dog frame) it follows directly that

dogof (J.S., Fido) (4)

whence, by the criteriality of Dogowner, from (1) and (4), we have:

Dogowner (J.S.)

The translation of this piece of reasoning into the functional notation is left as an exercise for the reader.

All the examples of 'matching' I have seen have this rather simple character. More profound examples are hinted at in (Bobrow/Winograd, 1977b), however. *So far as one can tell, the processes of reasoning involved may be expressible only in higher-order logic. For example, it may be necessary to construct new relations by abstraction during the "matching" process. It is known (Huet, 1972; Pietrzykowski/Jensen, 1973) that the search spaces which this gives rise to are of gr[illegible] complexity, and it is not entirely clear that it will be possible to automate this process in a reasonable way.)*

This reading of a frame as an assertion has the merit of putting frames, frame-instances and 'matching' assumptions into a common language with a clear extensional semantics which makes it quite clear what all these structures *mean*. The (usual) inference rules are clearly correct, and are sufficient to account for most of the deductive properties of frames which are required. Notice, for example, that no special mechanism is required in order to see that J.S. is a Dogowner: it follows by ordinary first-order reasoning.

One technicality is worth mentioning. In KRL, the same slot-name can be used in different frames to mean different relations. For example, the *age* of a person is a number, but his *age* as an airline passenger (i.e. in the traveller frame) is one of {infant, child, adult}. We could not allow this conflation, and would have to use different names for the different relations. It is an interesting exercise to extend the usual first-order syntax with a notion of name-scope in order to allow such pleasantries. But this is really nothing more than syntactic sugar.

Seeing As

One apparently central intuition behind frames, which seems perhaps to be missing from the above account, is the idea of *seeing* one thing *as though* it were another: or of specifying an object by comparison with a known prototype, noting the similarities and points of difference (Bobrow/Winograd, 1977a). This is the basic analogical reasoning behind MERLIN (Moore/Newell, 1973), which Minsky cites as a major influence.

Now this idea can be taken to mean several rather different things. Some of them can be easily expressed in deductive-assertional terms, others less easily.

The first and simplest interpretation is that the 'comparison' is filling-in the details. Thus, to say JS is a man tells us something about him, but to say he is a bus conductor tells us more. The bus conductor frame would presumably have slots which did not appear in the Man frame (*since-when* for example, and *bus-company*), but it would also have a slot to be filled by the Man instance for JS (or refer to him in some other way), so have access to all his slots. Now there is nothing remarkable here. All this involves is asserting more and more restrictive properties of an entity. This can all be done within the logical framework of the last section.

The second interpretation is that a frame represents a 'way of looking' at an entity, and this is a *correct* way of looking at it. For example a Man may also

be a Dog-owner, and neither of these is a *further* specification of the other: each has slots not possessed by the other frame. Thus far, there is nothing here more remarkable than the fact that several properties may be true of a single entity. Something may be both a Man *and* a Dog-owner, of course: or both a friend *and* an employee, or both a day *and* a birthday. And each of these pairs can have its own independent criteriality.

However, there is an apparent difficulty. A single thing may have apparently contradictory properties, seen from different points of view. Thus, a man viewed as a working colleague may be suspicious and short tempered; but viewed as a family man, may have a sweet and kindly disposition. One's views of oneself often seem to change depending on how one perceives one's social role, for another example. And in neither case, one feels, is there an outright contradiction: the different viewpoints 'insulate' the parts of the potential contradiction from one another.

I think there are three possible interpretations of this, all expressible in assertional terms. The first is that one is really asserting different properties in the two frames: that 'friendly' at work and 'friendly' at home are just different notions. This is analogous to the case discussed above where 'age' means different relations in two different contexts. The second is that the two frames somehow encode an extra parameter: the time or place, for example: so that Bill really is unfriendly *at work* and friendly *at home*. In expressing the relevant properties as assertions one would be obliged then to explicitly represent these parameters as extra arguments in the relevant relations, and provide an appropriate theory of the times, places, etc. which distinguish the various frames. These may be subtle distinctions, as in the self seen-as-spouse or the self seen-as-hospital-patient or seen-as-father, etc., where the relevant parameter is something like interpersonal role. I am not suggesting that I have any idea what a theory of these would be like, only that to introduce such distinctions, in frames or any other formalism, is to assume that there *is* such a theory-perhaps a very simple one. The third interpretation is that, after all, the two frames contradict one another. Then of course a faithful translation into assertions will also contain an explicit contradiction.

The assertional language makes these alternatives explicit, and forces one who uses it to choose which interpretation he means. And one can always express that interpretation in logic. At worst, *every* slot-relation can have the name of its frame as an extra parameter, if really necessary.

There is however a third, more radical, way to understand seeing-as. This is to view a seeing-as as a metaphor or analogy, without actually asserting that it is *true*. This is the MERLIN idea. Example: a man may be looked at as a pig, if you think of his home as a sty, his nose as a snout, and his feet as trotters. Now such a caricature may be useful in reasoning, without its being taken to be veridically true. One may *think of* a man as a pig, knowing perfectly well that as a matter of fact he isn't one.

MERLIN's notation and inference machinery for handling such analogies are very similar respectively to frames and "matching", and we have seen that this is merely first-order reasoning. The snag is that we have no way to distinguish a 'frame' representing a mere caricature from one representing a real assertion. Neither the old MERLIN (in which *all* reasoning is this analogical reasoning) nor KRL provide any means of making this rather important distinction.

What does it *mean* to say that you can look at a man as a pig? I think the only reasonable answer is something like: certain of the properties of (some) men are preserved under the mapping defined by the analogy. Thus, perhaps, pigs are greedy, illmannered and dirty, their snouts are short, upturned and blunt, and they are rotund and short-legged. Hence, a man with these qualities (under the mapping which defines the analogy: hence, the man's *nose* will be upturned, his *house* will be dirty) may be plausibly be regarded as pig-like. But of course there are many other properties of pigs which we would *not* intend to transfer to a men under the analogy: quadrupedal gait, being a source of bacon, etc. (Although one of the joys of using such analogies is finding ways of extending them: "Look at all the little piggies ... sitting down to eat their bacon" [G. Harrison]). So, the intention of such a caricature is, that some -not all- of the properties of the caricature shall be transferred to the caricaturee. And the analogy is correct, or plausible, when these transferred properties do, in fact, hold of the thing caricatured: when the man *is* in fact greedy, slovenly, etc....

This is almost exactly what the second sense of seeing-as seemed to mean: that the man 'matches' the pig frame. The difference (apart from the systematic rewriting) is that here we simply cannot assume criteriality of this pig frame. To say that a man *is* a pig is false: yet we have assumed that this fellow does fit this pig frame. Hence the properties expressed in this pig frame cannot be criterial for pig. To say that a man *is* a pig is to use criteriality incorrectly.

This then helps to distinguish this third sense of seeing-as from the earlier senses: the failure of criteriality. And this clearly indicates why MERLIN and KRL cannot distinguish caricatures from factual assertions; for criteriality is not made explicit in these languages. We can however easily express a non-criterial frame as a simple assertion.

One might wonder what use the 'frame' idea is when criteriality is abandoned, since a frame is now merely a conjunction. Its boundaries appear arbitrary: why conjoin just these properties together? The answer lies in the fact that not *all* properties of the caricature are asserted of the caricaturee, just those bundled together in the seeing-as frame. The bundling here is used to delimit the scope of the transfer. We could say that these properties were criterial for *pig-likeness* (rather than *pig-hood*).

In order to express caricatures in logic, then, we need only to define the systematic translations of vocabulary: nose — snout, etc., this seems to require some syntactic machinery which logic does not provide: the ability to substitute one relation symbol for another in an assertion. This kind of "analogy map-

ping" was first developed some years ago by R. Kling and used by him to express analogies in mathematics. Let ϕ be the syntactic mapping 'out' of the analogy (e.g. ⌜snout⌝ → ⌜nose⌝, ⌜sty⌝ → ⌜house⌝), and suppose $\lambda x.\ \psi(x)$ is the defining conjunction of the frame of Pig-likeness:

$$\text{Pig-like}(x) \equiv \psi(x)$$

(Where ψ may contain several existentially bound variables, and generally may be a complicated assertion). Then we can say that Pig-like (Fred) is true just when $\phi(\psi)$ holds for Fred, i.e. the asserted properties are *actually* true of Fred, when the relation names are altered according to the syntactic mapping ϕ. So, a caricature frame needs to contain, or be somehow associated with, a specification of how its vocabulary should be altered to fit reality. With this modification, all the rest of the reasoning involved is first-order and conventional.

Defaults

One aspect of frame reasoning which is often considered to lie outside of logic is the idea of a default value: a value which is taken to be the slot filler in the absence of explicit information to the contrary. Thus, the default for the *home-port* slot in a traveller frame may be the city where the travel agency is located (Bobrow et al. 1977).

Now, defaults certainly seem to take us outside first-order reasoning, in the sense that we cannot express the assumption of the default value as a simple first-order consequence of there being no contrary information. For if we could, the resulting inference would have the property that $p \vdash q$ but $(p \,\&\, r) \vdash \neg q$ for suitable p, q and r (p does not deny the default: q represents the default assumption: r overrides the default), and no logical system behaves this way (Curry [1956] for example, takes $p \vdash q \Rightarrow p \,\&\, r \vdash q$ to be the fundamental property of *all* 'logistic' systems).

This shows however only that a *naive* mapping of default reasoning into assertional reasoning fails. The moral is to distrust naivety. Let us take an example. Suppose we have a Car frame and an instance of it for my car, and suppose it has a slot called *status*, with possible values {*OK*, *struggling*, *needs-attention*, *broken*}, and the default is OK. That is, in the absence of contrary information, I assume the car is OK. Now I go to the car, and I see that the tyre is flat: I am surprised, and I conclude that (contrary to what I expected), the correct filler for the *status* slot is *broken*. But, it is important to note, my state of knowledge has changed. I was previously making an assumption — that the car was OK — which was reasonable *given my state of knowledge at the time*. We might say that if ψ represented my state of knowledge, then status (car) = OK was a reasonable inference from ψ: $\psi \vdash$ status (car) = OK. But once I know the tyre is flat, we have a new state of knowledge ψ_1, and of course $\psi_1 \vdash$ status (car) = broken. *In order for this to be deductively possible, it must be that ψ_1 is got from ψ not merely by adding new beliefs, but also by removing some old ones. That is, when I see the flat tyre I am surprised:* I had expected that it was OK. (This is not to say that I had explicitly considered the possiblity that the tyre might be flat, and rejected it. It only means that my state of belief was such that the tyres being OK was a consequence of it). And of course this makes sense: indeed, I was surprised. Moreover, there is no contradiction between my earlier belief that the car was OK and my present belief that it is broken. If challenged, I would not say that I had previously been irrational or mad, only misinformed (or perhaps just *wrong*, in the sense that I was entertaining a false belief).

As this example illustrates, default assumptions involve an implicit reference to the whole state of knowledge at the time the assumption was generated. Any event which alters the state of knowledge is liable therefore to upset these assumptions. If we represent these references to knowledge states explicitly, then 'default' reasoning can be easily and naturally expressed in logic. To say that the default for *home-port* is Palo Alto is to say that unless the current knowledge-state says otherwise, then we will assume that it is Palo Alto, *until the knowledge-state changes*. Let us suppose we can somehow refer to the current knowledge-state (denoted by NOW), and to a notion of derivability (denoted by the turnstile $\vdash$). Then we can express the default assumption by:

$\exists y.\ \text{NOW} \vdash$ ⌜*homeport* (traveller) = y⌝ ∨ *homeport* (traveller) = Palo Alto. The conclusion of which allows us to infer that *homeport* (traveller) = Palo-Alto *until the state of knowledge changes*. When it does, we would have to establish this conclusion for the new knowledge state.

I believe this is intuitively plausible. Experience with manipulating collections of beliefs should dispel the feeling that one can predict all the ways new knowledge can affect previously held beliefs. We do not have a theory of this process, nor am I claiming that this notation provides one.* But *any* mechanism — whether expressed in frames or otherwise — which makes strong assumptions on weak evidence needs to have some method for unpicking these assumptions when things go wrong, or equivalently of controlling the propagation of inferences from the assumptions. This inclusion of a reference to the knowledge-state which produced the assumption is in the latter category. An example of the kind of axiom which might form part of such a theory of assumption-transfer is this. Suppose $\phi \vdash p$, and hence p, is in the knowledge-state ϕ, and suppose we wish to generate a new knowledge-state ϕ' by adding the observation q. Let ψ be $\phi -$ ⌜$\phi \vdash p$⌝ and all inferred consequences of ⌜$\phi \vdash p$⌝. Then if $\psi \cup \{q\} \nvdash \neg p$, define ϕ' to be $\psi \cup$ ⌜$\psi \vdash p$⌝,⌜q⌝. This can all be written, albeit rather rebarbitively, in logic augmented with notations for

* Recent work of Doyle, McDermott and Reiter is providing such a theory: see (Doyle, 1978) (McDermott/Doyle, 1978) (Reiter, 1978)

describing constructive operations upon knowledge-states. It would justify for example the transfer of *status* (car) = OK past an observation of the form, say, that the car was parked in an unusual position, provided that the belief state did not contain anything which allowed one to conclude that an unusual parking position entailed anything wrong with the car. (It would also justify transferring it past an observation like *it is raining*, or *my mother is feeling ill*, but these transfers can be justified by a much simpler rule: if p and q have no possible inferential connections in ϕ — this can be detected very rapidly from the 'connection graph' (Kowalski 1973) — then addition of q cannot affect p.)

To sum up, a close analysis of what defaults mean shows that they are intimately connected with the idea of *observations*: additions of fresh knowledge into a data-base. Their role in *inference* — the drawing of consequences of assumptions — is readily expressible in logic, but their interaction with observation requires that the role of the state of the system's own knowledge is made explicit. This requires not a new *logic*, but an unusual *ontology*, and some new primitive relations. We need to be able to talk *about the system itself*, in its own language, and to involve assumptions about itself in its own processes of reasoning.

Reflexive Reasoning

We have seen that most of 'frames' is just a new syntax for parts of first-order logic. There are one or two apparently minor details which give a lot of trouble, however, especially defaults. There are two points worth making about this. The first is, that I believe that this complexity, revealed by the attempt to formulate these ideas in logic, is not an artefact of the translation but is intrinsic to the ideas involved. Defaults just *are* a complicated notion, with far-reaching consequences for the whole process of inference-making. The second point is a deeper one.

In both cases — caricatures and defaults — the necessary enrichment of logic involved adding the ability to talk about the system itself, rather than about the worlds of men, pigs and travel agents. I believe these are merely two relatively minor aspects of this most important fact: much common-sense reasoning involves the reasoner in thinking about himself and his own abilities as well as about the world. In trying to formalise intuitive common-sense reasoning I find again and again that this awareness of one's own internal processes of deduction and memory is crucial to even quite mundane arguments. There is only space for one example.

I was once talking to a Texan about television. This person, it was clear, knew far more about electronics than I did. We were discussing the number of lines per screen in different countries. One part of the conversation went like this.

Texan: You have 900 lines in England, don't you?
Me: No, 625.
Texan (confidently): I *thought* it was 900.
Me (somewhat doubtfully): No, I think it's 625.
(pause)
Say, they couldn't change it without altering the sets, could they? I mean by sending some kind of signal from the transmitter or ....
Texan: No, they'd sure have to alter the receivers.
Me (now confident): Oh, well, it's definitely 625 lines then.

I made a note of my own thought processes immediately afterwards, and they went like this. I *remembered* that we had 625 lines in England. (This remembering cannot be introspectively examined: it *seems* like a primitive ability, analogous to FETCH in CONNIVER. I will take it to be such a primitive in what follows. Although this seems a ludicrously naive assumption, the internal structure of remembering will not concern us here, so we might as well take it to be primitive.) However, the Texan's confidence shook me, and I examined the belief in a little more detail. Many facts emerged: I remembered in particular that we had changed from 405 lines to 625 lines, and that this change was a long, expensive and complicated process. For several years one could buy dual-standard sets which worked on either system. My parents, indeed, had owned such a set, and it was prone to unreliability, having a huge multigang sliding-contact switch: I had examined its insides once. There had been newspaper articles about it, technical debates in the popular science press, etc.. It was not the kind of event which could have passed unnoticed. (It was this *richness of detail*, I think, which gave the memory its subjective confidence: I couldn't have imagined all *that*, surely?) So if there had been another, subsequent, alteration to 900 lines, there would have been another huge fuss. But I had no memory at all of any such fuss: so it couldn't have happened. (I had a definite subjective impression of *searching* for such a memory. For example, I briefly considered the possibility that it had happened while my family and I were in California for 4 months, being somehow managed with great alacrity that time: but rejected this when I realised that our own set still worked, unchanged, on our return). Notice how this conclusion was obtained. It was the kind of event I would remember; but I don't remember it; so it didn't happen. This argument crucially involves an explicit assertion about my own memory. It is not enough that I didn't remember the event: I had to *realise* that I didn't remember it, and *use* that realisation in an argument.

The Texan's confidence still shook me somewhat, and I found a possible flaw in my argument. *Maybe* the new TV sets were constructed in a new sophisticated way which made it possible to alter the number of lines by remote control, say, by a signal from the transmitter. (This seems quite implausible to me now; but my knowledge of electronics is not rapidly accessible, and it did seem a viable possibility at the moment). How to check whether this was

possible? Why, ask the expert: which I did, and his answer sealed the only hole I could find in the argument.

This process involves taking a previously constructed argument — a proof, or derivation — as an object, and inferring properties of it: that a certain step in it is weak (can be denied on moderately plausible assumption), for example. Again, this is an example of *reflexive reasoning:* reasoning involving descriptions of the self.

Conclusion

I believe that an emphasis on the analysis of such processes of reflexive reasoning is one of the few positive suggestions which the 'frames' movement has produced. Apart from this, there are no new insights to be had there: no new processes of reasoning, no advance in expressive power.

Nevertheless, as an historical fact, 'frames' have been extraordinarily influential. Perhaps this is in part because the original idea was interesting, but vague enough to leave scope for creative imagination. But a more serious suggestion is that the *real* force of the frames idea was not at the representational level at all, but rather at the implementation level: a suggestion about how to organise large memories. Looked at in this light, we could sum up 'frames' as the suggestion that we should *store* assertions in nameable 'bundles' which can be retrieved via some kind of indexing mechanism on their names. In fact, the suggestion that we should store assertions in non-clausal form.

Acknowledgements

I would like to thank Frank Brown and Terry Winograd for helpful comments on an earlier draft of this paper.

Appendix: Translation of KRL-ϕ into Predicate Logic

KRL	*many-sorted predicate logic*
Units	
(i) Basic	Unary predicate (sort predicate: assuming a disjoint sort structure.)
(ii) Specialisation	Unary predicate
(iii) Abstract	Unary predicate
(iv) Individual	name (individual constant)
(v) Manifestation	sometimes a λ-expression λx. P(x) & ... & Q(x) sometimes an ∈-expression ∈x. P(x) & ... & Q(x) (i.e. a variable over the set {x: P(x) & ... & Q(x)}
(vi) Relation	relation
Slot	*binary relation or unary function*
Descriptors	
(i) direct pointer	name
(ii) Perspective e.g. (a trip with destination = Boston airline = TWA)	λ-expression e.g. λx. trip (x) & destination (x) = Boston & airline (x) = TWA (in this case both fillers are unique. If not we would use a relation, e.g. airline (x, TWA))
(iii) Specification e.g. (the actor from Act E17 (a chase ...))	ι-expression e.g. ιx. actor (E17) = x or ιx. actor (E17) = x & Act (E17)
(iv) predication	λ-expression
(v) logical boolean	non-atomic expression
(vi) restriction e.g. (the one (a mouse) (which owns (a dog)))	ι-expression e.g. ιx. mouse (x) & ∃y.dog (y) & owns (x, y)
(vii) selection e.g. (using (the age from Person this one) select from (which is less than 2) ~ Infant (which is at least 12) ~ Adult otherwise child	ι-expression with conditional body e.g. ιx. (age (this one) < 2 & x = infant) ∨ (age) (this one) ≥ 12 & x = adult) ∨ (age (this one) < 2 & age (this one) ≥ 12 & x = child)
(viii) set specification	λ-expression (sets coded as predicates) or set specification (if we use set theory. Only very simple set theory is necessary)
(ix) contingency e.g. (during state 24 then (the topblock from (a stack with height = 3)))	ι-expression or ∈-expression whose body mentions a state or has a bound state variable. e.g. ιx.∃y. is stack (y, state 24) & height (y) = 3 & topblock (y, x) where I have taken stack to be a contingent property: other choices are possible (e.g. stacks always "exist" but have zero height in some states).

Examples

Traveller (x) ⊃ Person (x) &
(category (x) = infant
∨ category (x) = child
∨ category (x) = adult)
& ∃y. airport (y) & preferredairport (x, y)

Person (x) ⊃ string (first name (x)) & string (last name (x))
& integer (age (x))
& city (nametown (x))
& address (streetaddress (x))

Person (G0043)
  & firstname (G0043) = "Juan"
  & foreignname (lastname (G0043))
  & firstcharacter (lastname (G0043)) = "M"
  & age (G0043) > 21
Traveller (G0043)
  & category (G0043) = Adult
  & preferredairport (G0043, SJO)

References

Bobrow, D.G., Kaplan, R.M., Norman, D.A., Thompson, H. and Winograd, T.
1977
"GUS, a Frame-Driven Dialog System", *Artificial Intelligence* 8, 155–173.

Bobrow, D.G. and Winograd, T.
1977a
"An Overview of KRL", *Cognitice Science* 1, 3–46.
1977b
"Experience with KRL-O: One Cycle of a Knowledge Representation Language", Proc. 5th Int. Joint Conf. on AI, MIT, (vol 1), 213–222.

Brady, J.M. and Wielinga, B.J.
1977
"Reading the Writing on the Wall", Proc. Workshop on Computer Vision, Amherst Mass.

Charniak, E.
1977
"Ms. Malaprop, a Language Comprehension Program", Proc. 5th Int. Joint Conf. on AI, MIT, (vol 1), 1–8.

Curry, H.B.
1956
Introduction to Mathematical Logic (Amsterdam: Van Nostrand)

Doyle, J.
1978
Truth Maintenance System for Problem Solving, Memo TR-419, A.I. Laboratory, MIT

Feldman, J.
1975
"Bad-Mouthing Frames", Proc. Conf. on Theor. Issues in Natural Language Processing", Cambridge Mass, 102–103.

Hayes, P.J.
1975
"Some Problems and Non-problems in Representation Theory", Proc. 1st AISB Conf., Brighton Sussex.
1977
"In Defence of Logic", 5 Int. Joint Conf. on AI, MIT, (vol 2), 559–565.

Huet, G.P.
1972
Constrained Resolution: a Complete Method for Type Theory, Jenning's Computer Science, Report 1117, Cace Western University.

Kowalski, R.
1973
An Improved Theorem-Proving System for First Order Logic, DCL Memo 65, Edinburgh.

Lehnert, W.
1977
"Human and Computational Question Answering", *Cognitive Science* 1, 47–73.

McCarthy, J. and Hayes, J.P.
1969
"Some Philosophical Problems from the Standpoint of Artificial Intelligence", *Machine Intelligence* 4, 463–502.

McDermott, D. and Doyle, J.
1978
Non-monotonic logic I, Memo AI-486, A.I. Laboratory, MIT

Minsky, M.
1975
"A Framework for Representing Knowledge", in P. Winston (Ed.) *The Psychology of Computer Vision*, (New York: McGraw-Hill), 211–277.

Moore, J. and Newell, A.
1973
"How Can MERLIN Understand?", in L. Gregg (Ed.) *Knowledge and Cognition* (Hillsdale New York: Lawrence Erlbaum Assoc), 201–310.

Pietrzykowski, T. and Jensen, D.
1973
Mechanising W-Order Type Theory through Unification, Dept. of Applied Analysis and Comp. Science, Report CS-73-16, University of Waterloo.

Reiter, R.
1978
"On Reasoning by Default", Proc. 2nd Symp. on Theor. Issues in Natural Language Processing, Urbana, Illinois.

Schank, R.
1975
"The Structure of Episodes in Memory", in D.G. Bobrow and A. Collins (Eds) *Representation and Understanding*, (New York: Academic Press), 237–272.

Wilks, Y.
1976
"Natural Language Understanding Systems within the AI Paradigm: a Survey", in M. Penny (Ed) *Artificial Intelligence and Language Comprehension*, (National Institute of Education, Washington, Oc).