A philosophical basis for knowledge acquisition

Garvan Institute of Medical Research technical report CS--89-01
(also CSIRO Division of Information Technology technical report TR-FD-89-01
and submitted to the 3rd european knowledge acquisition fo knowledge based systems workshop 1989)

P. Compton and R. Jansen†

Garvan Institute of Medical Research, St. Vincent's Hospital
Sydney 2010, Australia
†CSIRO Division of Information Technology
PO Box 1599 Macquarie Centre
Sydney 2113, Australia

Abstract

Knowledge acqusition for expert systems is a purely practical problem to be solved by experiment, independent of philosophy. However the experiments one chooses to conduct will be influenced by one's implict or explicit philosphy of knowledge, particularly if this philosophy is taken as axiomatic rather than as an hypothesis. We argue that practical experience of knowledge engineering, particularly in the long term maintenance of expert systems, suggests that knowledge does not necessarily have a rigorous structure built up from primitive concepts and their relationships. The knowledge engineer finds that the expert's knowledge is not so much recalled, but to a greater or lesser degree 'made up' by the expert as the occasion demands. The knowledge the expert provides varies with the context and gets its validity from its abiltity to explain data and justify the expert's judgement *in the context*. We argue that the physical symbol hypothesis with its implication that some underlying knowledge structure can be found is a misleading philosophical underpinning for knowledge acquisition and representation. We suggest that the 'insight' hypothesis of Lonergan better explains the flexibility and relativity of knowledge that the knolwedge engineer experiences and may provide a more suitable philosophical environment for developing knowledge acquisition and representation tools. We outline the features desirable in tools based on this philosophy and the progress we have made towards developing such tools.

Introduction

We wish to consider the philosophical question of the truth value of knowledge. This is not a question regarding the truth of a particular piece of knowledge, or a question regarding the ability of an expert, but a question about knowledge in general. In what way is knowledge really a representation of reality, how good a representation is it, how good a representation can it be. These are obviously time honoured questions in philosophy and we do not presume to suggest new answers but we do suggest that a

what you are acquiring will influence your methods. Nor do we suggest that the problems we will identify are unknown to the artificial intelligence community, in fact there are many many practical attempts to deal with various aspects of these problems. We suggest however a better epistemology may allow better solutions to be developed.

There has been some consideration of epistemology in recent years by workers such as the Dreyfus brothers (1988) who attack the whole artificial intelligence (AI) venture, or others such as Winograd and Flores (1987) who suggest that AI's real achievements will be somewhat different from the achievement of machine intelligence. The line of the Dreyfus attack is that since artificial intelligence makes incorrect assumptions about the nature of knowledge it cannot succeed. One response of the AI community is that since AI does work and we can build expert systems, then the philosophy must be correct. Our aim here is to avoid this polarisation by focussing on a replacement or modification rather than a rejection of what we perceive as the current AI philosophy of knowledge.

The prevailing epistemology is the physical symbol hypothesis of Newell and Simon (1981). Essentially this hypothesis considers that knowledge consists of symbols of reality and relationships between these symbols and that intelligence is the appropriate logical manipulation of the symbols and their relations. Although Newell an Simon developed these ideas in the 50's as a basis for AI, they have a long and time honoured philosophical ancestry, through the early Wittgenstein, back through Descartes to Plato with his archetypes.

AI research then is based on the assumption that one should be able to obtain in some way, fundamental atoms of knowledge and the logical relations between these atoms, and from these reassemble knowledge; it has an essentially reductionist strategy. The extension of the physical symbol hypothesis is the "knowledge principle", that is, the success of an expert system does not depend on the sophistication of its inferencing or reasoning strategy, but on the amount of information it contains on how symbols are interrelated, that is the amount of knowledge it contains. (Feigenbaum 1977). The fullest expression of this is the "knowledge is all there is" hypothesis (Lenat and Feigenbaum 1988), that is, there are no control or inferencing strategies that cannot themselves be contained in knowledge. Lenat (1988) would suggest that the CYC project, aimed at capturing much of common sense, or consensus reality, in a knowledge base is proceeding faster and faster because a fair number of the primitives out of which more complex knowledge is built have now been captured. Lenat would imply that the increasing success of the CYC project is a verification of the underlying philosophy.

Because the physical symbol hypothesis works so well it has become not one philosophy amongst others but an absolute viewpoint for AI workers. Thus the conventional explanation of the difficulty in acquiring knowledge is that experts don't communicate the underlying knowledge very well; e.g. as Waterman puts it ". . . . . the pieces of basic knowledge are assumed and combined so quickly that it is difficult for him *(the expert)* to describe the process." (Waterman 1986). With this type of perception, the problems in knowledge acquisition become, and have been over the last years, How do we really get to the bottom of the expert's knowledge, how do we really get down to the basic bits and their relationships, which can then be rebuilt into a

Bradsaw 1987) have been developed and some of these are available commercially (NEXTRA from Neuron Data (Rappaport and Gaines 1988)) which facilitate getting to the "bottom" of knowledge, and these have contributed and will contribute greatly to expert system development.

However there are clearly some problems. Shaw (1988) notes that contrary to the expectation of a knowledge engineer, different experts not only talk about a common topic of expertise in quite different terms, but are quite happy to disagree on terminology, without necessarily making this explicit. They use the same terminology in conceptually different ways and would appear to have different knowledge structures. Although these findings can of course be made to fit with the physical symbol hypothesis, they at least raise the question of whether there may be a better philosophy of knowledge than one whereby knowledge is made up from some sort of absolute primitive elements. Wittgenstein, in his later years, gave up logical atomism precisely because of this sort of failure to find the primitive concepts on which knowledge is built. (Wittgenstein 1953).

A different type of question is, What happened to all the prototype expert system? Clearly there have been far more systems prototyped than deployed? The phase of prototyping an expert system is the hopeful phase. You have got 90% of the knowledge into the system, and you are confident that if only you persevere, you will overcome the problems and the system's knowledge will become complete. Generally there are no further reports beyond the prototype stage, but where there are, a different story emerges. McNamara et al (1988) have started again with a new design to build an expert system which they were confident was close to completion in 1987 (Lock Lee et al 1987). After four years in routine use, knowledge addition to XCON/R1 still involved 4 full time knowledge engineers, although this was at least partly because the domain itself was expanding with new computer hardware to be configured (Bachant and MacDermott 1984). More recently a reimplementation of XCON has been commenced in an attempt to make it more maintainable (Soloway et al 1987). GARVAN-ES1 has doubled in size since 1984 when it was put into routine use although the knowledge domain has not expanded and the system was 96% correct when deployed (Horn et al. 1985 and Compton et al. 1988). We will argue below that such problems are not the brittleness, or ultimate stupidity, of expert systems due to lack of common sense that Lenat would fix with CYC, but problems related to what knowledge is.

We can also note that the apparent increased deployment of expert systems at present is somewhat misleading as an indicator of improvement in expert system building. There seems to be a fairly major shift in the type of expert systems being built. In the past medical diagnosis was the expert system paradigm, whereas now the focus is on areas such as finance and process control. These are quite different types of problems with quite different requirements of expertise. In medicine the starting point is that each individual patient should have the very best care and diagnosis available. In practice we then step back from this goal because of costs, human frailty, organisational problems and so on. An expert system in this environment must be 'truly expert', that is it must make no mistakes that could be picked up by a human. Schwartz et al (1987) on behalf of the medical AI community have given up this goal until large advances in technology are made. By contrast, in areas such as finance it is sufficient that the errors an expert

system (or an expert) makes are outweighed by its successes, according to the relevant profit criterion. Any improvement above this minimum is pure profit.

We would suggest that material for philosophical reflection is less obvious in these "error tolerant" systems than in systems where total expertise is demanded. We will consider here one attempt to perfect an expert system.

The maintenance experience

GARVAN-ES1 is a small medical expert system which consists of 96 rules which provide simple reduction of numeric and string data to classes and 650 production rules (without disjunctions) comprising the knowledge (Horn et al 1985). Its knowledge is entirely heuristic and contains no deep knowledge of the domain. The system is used to provide clinical interpretations of data for reports from a laboratory which measures thyroid hormone levels. It is forward chaining only as all the data is available prior to running the expert system during report generation. The system is a heuristic classification system (Clancey 1985) and chooses from between 60 different interpretations and provides about 6,000 interpretations per year. The system has been in routine use since mid 1984 and was identified by Buchanan (1986) as one of the first four medical expert systems to reach routine use. Most importantly it has been continuously maintained since its introduction (Compton et al 1988). Maintenance is made possible because each report issued by the system is checked and signed by a human expert or referred to the knowledge engineer if the clinical interpretation is unsatisfactory. Although 96% percent of the reports were acceptable to experts when the system was introduced into routine use, the knowledge base has more than doubled in size over the four years (Fig 1). with acceptance of the reports now apparently 99.7% (see Compton et al 1988a,b for the significance of this figure).

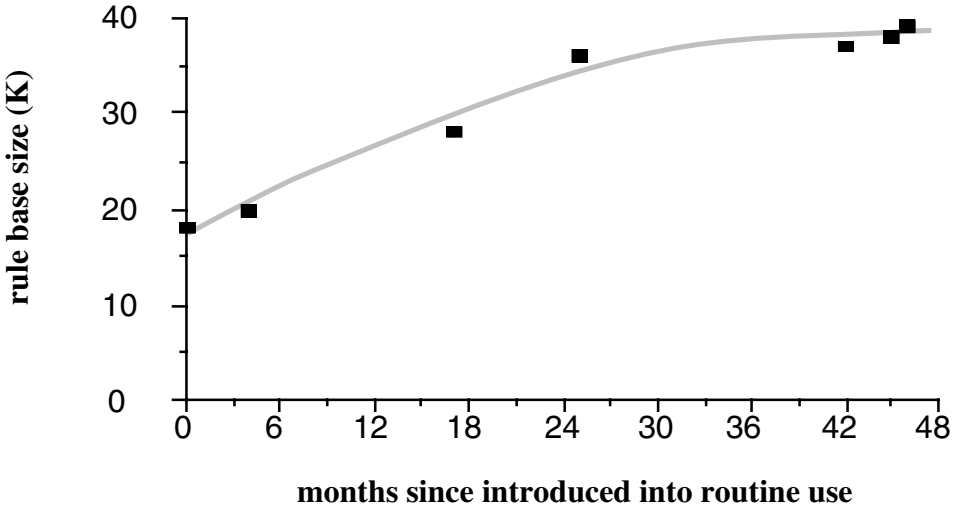

Fig 1. The growth of GARVAN-ES1 since its introduction into routine use in mid 1984. The Yaxis gives the number of characters in the rule base, and is best used to indicate the relative change in size of the knowledge base. The data points represent some copies of the rule base that were kept. Maintenance was carried out more regularly than these data points suggest. The curve was drawn by eye for illustrative purposes.

Compton et al (1988) provide an example of an individual rule that has increased four

or four different rules as the systems knowledge was refined. There are also examples, albeit very few, of later experts deciding that the knowledge contained in some early rules was completely spurious.

The most important resource for knowledge maintenance is a data base of "cornerstone cases". These are cases which at some stage have required a change in the system's knowledge. Every time the system's knowledge is changed the interpretations for all the cornerstone cases are checked to see that the additions to the system's knowledge have been incremental and have not corrupted the knowledge.

A basic experience of maintaining GARVAN-ES1, (and all knowledge engineering?) is that the knowledge one acquires from the expert has to be extensively manipulated so that it can be added to the knowledge base without corrupting the knowledge already contained. The difficulty of doing this is exacerbated during the maintenance phase not only because the knowledge base is becoming more complex, but because the expert and knowledge engineer are no longer closely familiar with the knowledge communicated during the prototype phase; in fact different experts and knowledge engineers may be involved, as has happened with GARVAN-ES1. Experts will still unfailingly give interpretations and good reasons for the interpretations for every set of data presented to them, but they are no longer at all familiar with the knowledge structure to which this knowledge is to be added. Shaw illustrates that experts have different knowledge structures concerning the same domain. (Shaw 88). We will argue below that even the knowledge provided by a single expert changes as the context in which this knowledge is required changes.

Figure 2 summarises the maintenance experience. The expert provides what appear to be very simple clear rules, however when these are added to the system they subsume and conflict with prexisting rules and must be extensively manipulated before they can be added to the knowledge base. This is not just a problem because the structure of a knowledge base is influenced by the inference strategy used on it (Bylander and Chandrasekaran 1986); the problem is in the knowledge provided by the experts. Further, the problem cannot be resolved by tools which check knowledge bases for subsumption etc (Suwa et al 1984) because rules which have no logical overlap may still be satisfied by a single data profile.

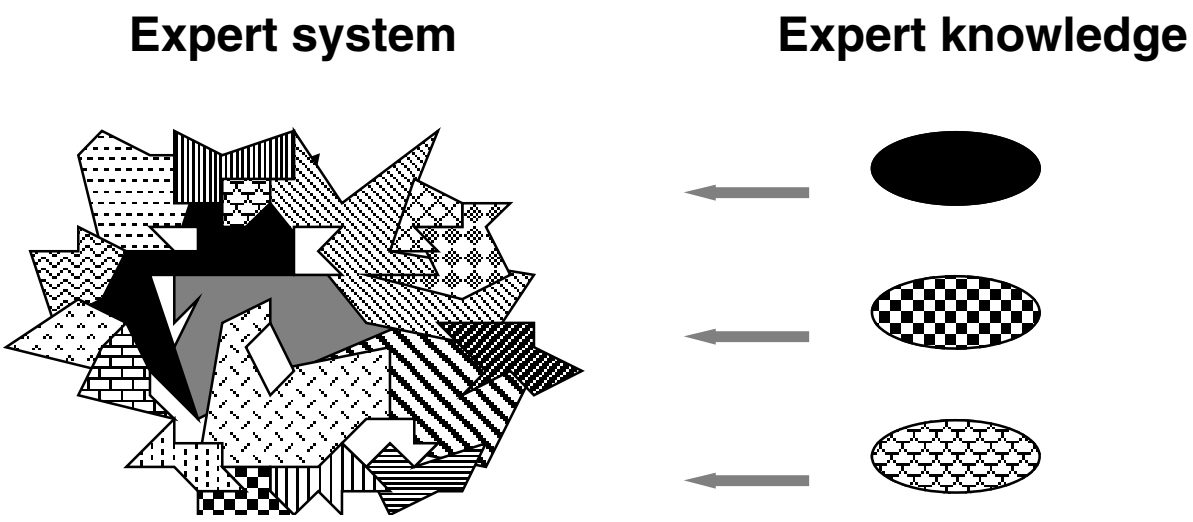

Fig 2. This illustrates the large difference between expert knowledge as expressed by the expert and as contained in an expert system. The diagram suggests how rules as expressed by experts must be manipulated and changed to ensure that they do not overlap with other rules. The physical symbol hypothesis would hold that the structure on the left is some sort of a map of knowledge of a domain.

Our previous study (Compton et al 1988) provides a detailed example of maintenance. This example illustrated that in the maintenance phase the context influences the rules that the expert provides, a very non controversial proposition. For example if a set of data was given interpretation **B** by the expert system when it should have been given interpretation **A**, then the rules the expert provides to correctly reach interpretation **A** will differ from those when the expert system had given interpretation **C** rather than **A**. No one would dispute that this will occur, that the context in which questions are asked will influence the answers. However the physical symbol hypothesis in fact suggests that something different should occur, that we should be able to obtain from experts the primitives and structure of their knowledge, that we should be able to obtain a rule to reach conclusion **A** instead of a group of more relative rules which conclude **A** rather than **B** and **A** rather than **C** etc. The physical symbol hypothesis suggests that there should be underlying rules that provide a more absolute foundation for the knowledge, rules to reach **A** rather than **not A**, i.e. context free rules. It has been harder than expected to find such foundations, hence the new knowledge acquisition tools (NEXTRA etc) aimed precisely at broadening the context to enable the expert to see all the bits and pieces of knowledge they provide, so that the structure can be found. These work, structures can be found, but they appear to differ from expert to expert (Shaw 1988).

The experience with GARVAN-ES1 suggests, rather than arriving at the ultimate structure, the knowledge acquisition process goes on and on, albeit more slowly; even the rule changes introduced during the maintenance example alluded to above and described previously (Compton et al 1988) have required further changes introduced during the writing of this manuscript. The rule base has doubled in size in trying to go from an accuracy of 96% to an accuracy of 99.7% (probably not as high as this in fact(Compton et al 88). When one examines the composition of GARVAN-ES1's data base of cornerstone cases, as a record of the growth of the system, one finds that rules for even the simplest and most clear cut interpretations are not fully correct when they were first entered, and it was necessary for the rules to misinterpret cases and fail to interpret cases a many times before they became reliable (Compton et al 1988). Experts however have little trouble in interpreting these cases and providing new rules; which they will then endlessly change when provided with other cases. It should be noted that the domain of interpretation of thyroid reports is a simple domain in medicine where you would expect it to be easy to get complete knowledge.

A philosophy of knowledge

This maintenance experience suggests that knowledge is not sometimes, but always given in context and so can only be relied on to be true in that context. The best philosophical foundation for this seems to be the philosophy of Karl Popper (Popper 1963). Popper suggests that we can never prove a hypothesis, we can only disprove hypotheses. Obviously we cannot disprove all alternate hypotheses, we may not even

number of *likely* alternatives. We suggest that this explains the phenomenon in knowledge aquisition that knowledge, the rules for reaching some conclusion, always seem to be context dependant. What the expert is doing is identifying features in the data that enable one to conclude that a certain interpretation *is preferable to the small set of other likely interpretations*, and of course the likely alternatives depend on the context. This is quite different from reporting on how one reached a given conclusion. These are not novel suggestions. It is well established in medicine for example, that clinicians use hypothetico deductive methods, that they attempt to distinguish between a small set of hypotheses (Griner 81). In earlier studies we have suggested that it is precisely the hypothetico deductive approach which allows clinicians to deal with the errors that occur in laboratory data. (Compton et al 86)

We would suggest that the knowledge that experts provide is essentially a justification of why they are right, not the reasons they reached this right conclusion. This phenomenon is familiar also in arguments. Arguments are always about showing how we are correct in terms that will convince the other person. No one is ever wrong but the context in which one is right changes during the argument! Similarly when an expert is explaining something, his explanation varies depending on whether he is explaining it to a layperson, a novice or a peer. If the expert was explaining, giving the "real" reasons, rather than justifying, they would provide the most detail to the layperson, who knows nothing about the area and the least to a peer, who knows a lot about the area. In fact the reverse happens, the layperson gets the most superficial explanation, because generally it requires very little to justify that the expert's hypothesis is better than the naive hypotheses of the lay person, while the fellow expert, precisely because he is an expert, will require more rigor in the justification. It has been easy to witness this phenomenon with GARVAN-ES1. If the knowledge engineer takes a difficult case to two experts independently, he will get two fairly simple, but sometimes slightly different rules, although each will apparently peform equally well in the expert system. If the knowledge engineer then brings the experts together and asks which rule is right, a very complex discussion is liable to ensue, as the experts (politely) attempt to prove to each other that their rule is better, normally resolving the question, by agreeing that their rules apply in different contexts and are complementary..

It seems to us that the proposition that the knowledge we communicate is a justification in a context, not how we reached our conclusion, is universally true of all knowledge. Popper, although his ideas are mainly applied to scientific knowledge, applied them to all knowledge (Popper 1963). A baby makes hypotheses about the world, which are then tested against experience and new hypotheses constructed, or the old ones modified, if the hypotheses are seen to fail. This would seem to fit closely with Piaget's observations of development in early childhood. (Piaget 1929). There have of course been additions to Popper's falsification hypothesis. Kuhn has drawn attention to paradigm shifts in science, where investigation shifts to new types of hypotheses for other reasons than the falsification of earlier hypotheses.(Kuhn 1962) Lakatos (1978) points out how falsification of hypotheses may result in riders being added to allow the major hypothesis to continue, although it becomes very unwieldy. An example would be Ptolemy's addition of epicycles, when it was found that planetary orbits were not circular, in order to maintain the primacy of circular motion. These developments do not disagree with the essential proposition of interest here, that the knowledge we communicate to each other is a justification of a judgement expressed in a specific

This viewpoint has a number of consequences. If all knowledge is only true in a context, then all knowledge is relative, it only exists in relation to other knowledge, there is no absolute underlying knowledge on which the rest of knowledge is built. We can note that AI does attempt to deal with the problems that flow from this. For example circumscription (McCarthy 1980) is an attempt to deal with the exceptions to commonsense beliefs, that always occur, but they are still seen as only exceptions, rather than as indicators of the fundamental relativity of all knowledge.

How then can knowledge be in any sense true, our original question? Before attempting to answer this we must consider the most common response to such suggestions; that is, that underlying primitive knowledge is made up from sense data. On the contrary, it is clear that what we perceive depends on the context in which the perception takes place. There are well known examples such as when one wears glasses which invert the view seen. After a while the inversion is corrected and the world is seen the right way up (Stratton 1897). A small model of a distorted trapezoidal shaped room can be built, but with viewing apertures with the appropriate prisms so that the room appears of appropriate proportions when viewed through the prisms; the model table, windows doors etc all have their distortion corrected. If one is asked to touch various parts of the model with a stick poked in from above, it is difficult because one has a misleading view of the room. If one perseveres and learns to touch the different objects easily, the true distorted shape of the room is *seen* although one is still looking through the correcting prism.(Ames 1951, described by Bateson 1979) Clearly one's expectation determines what one sees, i.e one's intellectual context determines the so-called prime data of sense. "The Logic of Perception" by Lock (1980) is entirely devoted to demonstrating in a formal way that the mind determines what one is able to perceive. Colour perception provides another interesting example of the same phenomenon. There are a number of tribal groups throughout the world who perceive fewer colours than Western man (Treitel 1984). This is not due to any impairment of visual apparatus; Taylor (1984) records how as a chemistry teacher in Nigeria, he found it very difficult to teach the use of litmus paper, because students, saw red and blue litmus paper as the same colours. However after much practise, they were able to distinguish the two, that is they were able to see red and blue as different colours. McLuhan and Parker(1969) note that perspective, the vanishing viewpoint, was introduced into art at a specific stage of history. Art prior to this time showed great technical skill but no perspective. There was then a sudden change to use of perspective. Since the artists surely had the ability to paint perspective, but didn't, perhaps they didn't *see* things in perspective till perspective was discovered or created. In fact, McLuhan and Parker's comments are much stronger and they would argue that there has been a continuing change in our perception of space and time.

Sense data then cannot be viewed as providing the primitives out of which other knowledge is built, because the sense data available to us is determined by our mind, by the knowledge we have and the context that this provides for perception. It is possible to argue that the chemical changes resulting from perception may be accumulated and recorded below consciousness. Such data may exist but is clearly irrelevant if we are seeking for the expert's knowledge, that is what the expert is conscious of and can reason about. We can also note that an appeal to chemical storage of information as providing the primitives of knowledge is contradictory in that it is based on ideas of

Although we have not used their arguments to support our position we can also note the similarity of this viewpoint to some of modern physics, where the world that is discovered through experiment, is determined by the knowledge we already have. The concrete realness of space and time that common sense suggests, are our creation to express our experience, rather than some absolute of reality in itself. (d'Espagnat 1983 for a restrained exposition of these ideas).

The picture of knowledge that has emerged then is one where knowledge only has meaning in relation to other knowledge. Any set of apparent primitives that one may find only have their meaning in relation to other knowledge. It is interesting to note that the new knowledge elicitation techniques, based on psychological rather than philosphical theory, treat knowledge in this way and attempt to explore knowledge in terms of relationships. But we are not back to a version of the physical symbol hypothesis whereby the set of relationships of the whole knowledge structure provide the foundation for knowledge. On the contrary our argument has been that any knowledge structure that is elicited is itself always provided in a context and cannot be guaranteed not to change when the context changes, and we note again Shaw's observation of the different knowledge structures experts have of the same reality. (Shaw 88)

What then is knowledge, how can we communicate knowledge if it is all relative, what can it mean to judge that something is true? The most satisfactory theory seems to be that of Lonergan (1959) who in some 900 pages describes the concept of insight, and its role in the different kinds of knowledge. Essentially the act of insight is our act of recognition that something does make sense. The physical symbol hypothesis viewpoint sees only knowledge. Insight is the act whereby we perceive that this knowledge makes sense of reality, expresses some intelligibility in reality. Lonergan specifically chose the word 'insight' to express, the excitement, the flash of discovery. Archimedes "Eureka!" was because he discovered a theory *that made sense of reality*. Once the theory was expressed, one could examine its logical structure, and express it in terms of a physical symbols, relations and logic, but first of all a theory had to be created and the soundness of its semantics as well as its logic recognised in the act of insight. Some of the earlier discussion may have seemed to align us with Kant, and with his more idealist followers, that knowledge and sensation are something internal with a tenuous or non-existent connection to reality. Lonergan's approach is descended from the realism of the Scholastics, a realism divorced from the essentially naive realism of rationalism. We are directly and ontologically in touch with reality in insight, but our expression of this is intrinsically different from reality; knowledge and perception never contain what is known. We never fully grasp and express and contain in our knowledge the intelligibility of reality, which of course leads back to Popper's hypothesis. Knowledge never expresses reality but we progress in knowledge, through disproving hypotheses in the unending search to express insight truly. We use our knowledge to try and express insight, but it is also part of the insight process, because the building of the knowledge is part of the process of seeing the intelligibility in reality.

Communication of knowledge can then have a two fold function. We can use it as a justification, to show that out of the alternative hypotheses, ours is the best. Or we can use it to try and convey insight, to enable the other person to recognise, even partially

conversations are where people are trying to show each other they are right and the best kind are where they are genuinely trying to see and enable each other to see their respective insights. Insight can of course often be expressed very badly and rigorous debate, scientific or otherwise, is necessary to see whether the hypothesis used to express the insight is wrong on some other ground. This does not mean the insight is wrong, but the expression of the insight is contradicted by the knowledge which expresses other insights. Even if the way the new insight is expressed seems perfect, it is by definition an hypothesis to be eventually supplanted by a "less wrong" or perhaps more interesting hypothesis.

We do then build up a body of knowledge, a knowledge structure. The trouble with this structure, is that it does not get its value because it is assembled from more primitive, more true, elements. It get its value because the various hypotheses that express the various insights don't conflict, they are consistent and coherent. But they are only consistent and coherent, as far down as they have been checked, and they have only been checked as far as demanded by the interaction of the various hypotheses, and the insights they express. As these change and develop the knowledge structure changes. There is no doubt there is a knowledge structure, but it gets its validity from the consistency and coherence of the structure, not from the elements that make it up. The structure of course also gets its validity, from how well it expresses insight and retains contact with insight. If the expressed knowledge is taken as the truth, disconnected from insight, the body of knowledge rapidly becomes corrupt. The weird aberrations that arise in philosophies, religions, civilisations if they lose their grounding insights attest to this phenomenon.

We can also note that insight is not something that happens automatically and uniformly. An expert can probably be defined as someone whose technical knowledge cannot be faulted and whose decisions will meet with the approval of his peer group and those who would seek his opinion. But there are also 'experts' in every discipline who are lateral thinkers, who are able to propose solutions for completely novel situations. At the other end of the scale are those who have no insight and attempt to apply textbook reasoning with disasterous results. Lonergan treats in detail the factors involved in insight.

Implications for expert systems and knowledge acquisition.

Any system that attempts to reach the fundamentals of knowledge on which all the other knowledge can be built even in a specific area , will not work, or will work by good fortune rather than design. The design principle of the vast body of expert systems, that you should try to get the general rules, that knowledge should be separated from context and generalised, is philosophically unsound.

Experts will always give knowledge in context, with rules that conflict, subsume, overlap etc. If you push them, they can express a body of knowledge that is more complete, consistent and coherent. But you have to push them, they are not digging out some underlying knowledge structure, they are constructing something, making something up to satisfy the knowledge engineering demand. Hence Shaw's finding that the knowledge structures that experts made up differed; and she further notes that the experts were happy to live with these differences and didn't see it as a matter of concern.

It is not suprising that CYC is starting to assemble useful "primitives", and that analogy is an increasing powerful tool in its repetoire; we do all communicate, so we must have something in common. But our communication is based on being able constantly to revise the language we use to express insights. If CYC becomes a monolith, what will happen with paradigm shifts. Kuhn has noted how scientific hypotheses of interest change for other reasons than just falsification. So can common sense; new insights are always possible, and can change the whole emphasis on what are the important areas to be made consistent and coherent in knowledge. The function of the artist (and still the scientist?), is to open up new insights, to find new ways of viewing reality.

We would suggest then, that it is fundamental to the expert system enterprise, that knowledge NOT be generalised when it is acquired. It is fundamental to try and also record the context in which the knowledge is acquired, for it is in the context that it still has its closest link with insight, and therefore the most validity. The knowledge can always be generalised later. Inductive learning methods such as ID3 (Quinlan et al 1987) provide examples of generalising from the highly context based knowledge of individual cases. We are proposing here that even where knowledge is elicited from experts, it should be recorded in context. One attempt at doing this is described in Appendix A and Compton et al 1988. Context of course cannot be fully recorded for we cannot know everything that makes up the context, precisely because it is the context.

The second point is that since the knowledge base will have to change it must be designed for change. One should be able to investigate and interrogate the knowledge from any viewpoint that is desired and change any aspect of the knowledge without corrupting the whole, or at least be able to control the changes fully. Appendix 2 and Jansen and Compton 1988a &b, describe our progress towards this goal. The essence of our approach is to use data base technology, so that knowledge, eg rules, can be broken down into component parts and stored in relational form, with all the facilities for maintenance that this implies. In one sense the goal is maintenance, but maintenance is not an extra, a problem that has to be dealt with but is just an instance of the fundamentally fluid nature of knowledge.

The underlying problem is that computers are machines that manipulate symbols. The physical symbol hypothesis is perfect for describing computer intelligence. Computers demand a reductionist approach, where the whole, the body of knowledge is assembled from the parts. Unfortunately, human knowledge isn't made up of such primitives. It may look like it on some occasions, but essentially it is made up as occasion demands, and only within the context in which knowledge is provided, where insight is at work, can it really be guaranteed to have any consistency and coherence,.

This doesn't mean computers shouldn't be used for storing knowledge. On the contrary we suggest that they will be major tools for developing consistent knowledge bases. And perhaps one of their major roles in the area of knowledge and intelligence, will not be to provide expertise, but to enable humans to examine how a new expression of a new insight relates to the totality of what is "known". We see that perhaps the most powerful role for the computer in its present technological form is as a hypothesis tester rather than the dispenser of intelligence. We do not question, however that expert systems can and do work and have an important future and we hope that our own work

and use of computer knowledge bases, we will make more progress if we don't assume that the knowledge that is going into the computer has come from a computer like thing.

Appendix 1 - Ripple down rules

We have proposed elsewhere (Compton and Jansen 1988) a simple structure aimed at capturing, at least partially, the context in which knowledge is obtained from an expert. Maintenance incidents with GARVAN-ES1 always occur because the expert system misinterprets or fails to interpret a case. The context in which a new rule is obtained from an expert always includes the information that there was no interpretation made or a specific interpretation was wrongly selected by the expert system. In the case of the wrong interpretation, we use as the context the specific rule that gave the wrong interpretation. So the new rule that the expert provides only becomes a candidate to fire if the rule that previously fired incorrectly is again satisfied. We do this by including a LAST_FIRED(rule number) in the premise of the new rule, which must be satisfied as well as the other conditions.

All rules in this system include this LAST_FIRED(rule number) condition, but rules which are not added as corrections have LAST_FIRED(0) indicating that they are the initial candidates to fire when the data is first presented to the system. The rules are only passed through once, from the oldest to the newest, and of course once one of the LAST_FIRED(0) rules has fired none of the other LAST_FIRED(0) rules is eligible as now a specific rule has fired. Each rule added is put at the end of the list of rules. When an error occurs because the system has failed to make an interpretation the new rule becomes a new LAST_FIRED(0) rule. The implicit context when such a rule is obtained from an expert and added is that none of the knowledge in the system covers this case. The rule obtained from the expert will almost certainly subsume or overlap other rules in the knowledge base - this is the standard experience of knowledge engineering. With ripple down rules however, the rule can be as general as the expert likes, because it will only be tested after all the other knowledge in the knowledge base has been considered, exactly the situation in which it was acquired from the expert. The structure of ripple down rules can be described in a number of ways. Fig 3 shows an alternative description of the knowledge structure.

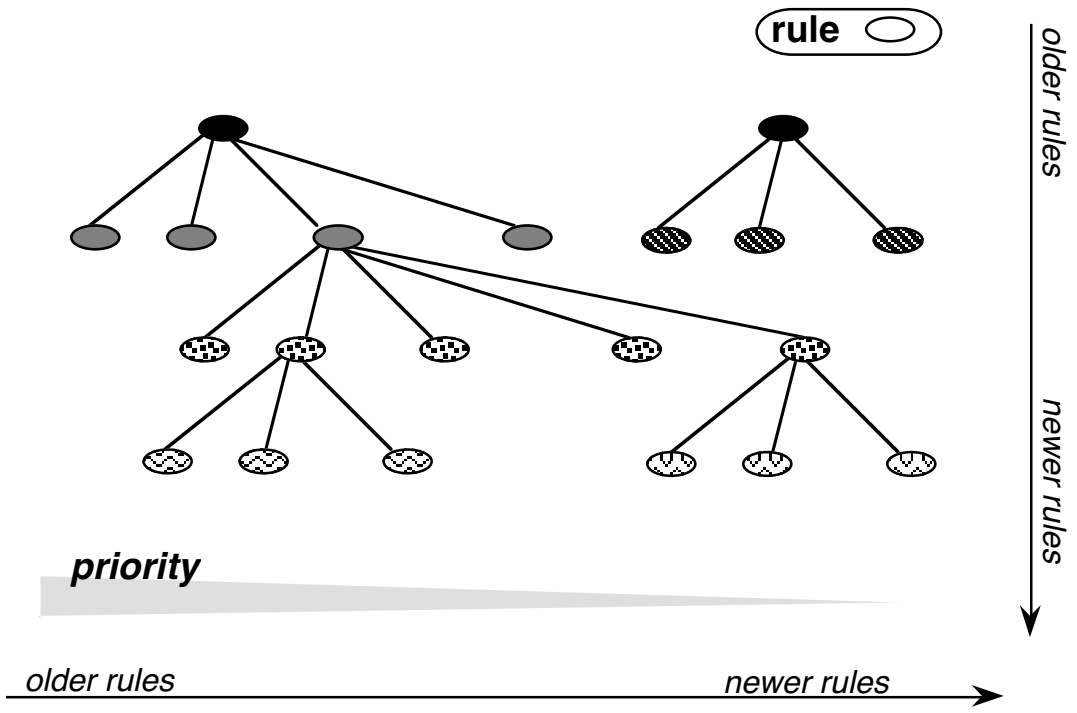

Fig 3. This is a simplified tree representation of ripple down rules. The two black elipses at the top of the tree are two LAST_FIRED(0) rules, so the precondition for either of these rules to fire is that no other rule has fired. The rule on the left, the earlier rule is considered first. If it doesn't fire, the next oldest LAST_FIRED(0) rule is considered. If a LAST_FIRED(0) rule does fire, then only the rules connected to it on the next level are candidates to fire. Again these are considered, one by one, left to right, the oldest to the newest added. Each of these has been added because the parent misinterpreted a case, but no further rule was able to fire. Each includes the LAST_FIRED(parent) condition. Once one of these daughter rules is added then the only the rules connected to it are candidates to fire.

This system means that rules can be entered into an expert system exactly as provided by the expert and the growth of the system's knowledge records the way in which knowledge was acquired from the expert. This is in complete contrast to conventional expert systems, where the aim is to generalise, and every rule (at the same level) has an equal opportunity to fire. However, the result of generalisation, is not clear general rules, but rules that become highly specific and opaque in trying to cope with the interrelations in knowledge in a general way.

In our previous study we described the incremental testing of, and addition to a set of ripple down rules to replace GARVAN-ES1, using a large data base of archived cases. This evaluation still is not complete but the essential behaviour of such rules is clear.

The major feature of ripple down rules is that they can be added to a knowledge base far faster than conventional rules since the rules are added as is without modification. Secondly, since they are used only in context they have far less impact in corrupting the knowledge base. (We note again that this is not a problem that can be resolved with tools that check for subsumption automatically etc (Suwa et al 1984), because often the overlap between rules is not in the rules themselves, but in the patient data profiles; the

adding 534 rules to the ripple down rule knowledge base, in only 31 cases did an extra rule have to be added because the first rule added had changed the performance of the knowledge base. This is in complete contrast to conventional rule addition. A log of recent maintenance activity on GARVAN-ES1 showed that it had taken 31 attempts to introduce 12 changes, with of course extensive checking of the knowledge base along the way. Because new rules have little effect on the ripple down rule knowledge base they can be added far faster than with a conventional system. It is not to difficult to add 10 rules per hour in contrast to the often mentioned industry figure of 2 rules per day. Such rules do work and Fig 4 illustrates the accuracy achieved in the evaluation so far. This evaluation is not complete but it seems likely that the new rule base will be as accurate as the original without being much larger than its 650 rules. With 40 fold faster rule addition, even a doubling in size would be insignificant. It will be necessary to retest the knowledge base on a further large data base of cases since as there are now only about 1500 cases that have not been used as the basis for rule changes and which can legitimately be used as test cases.

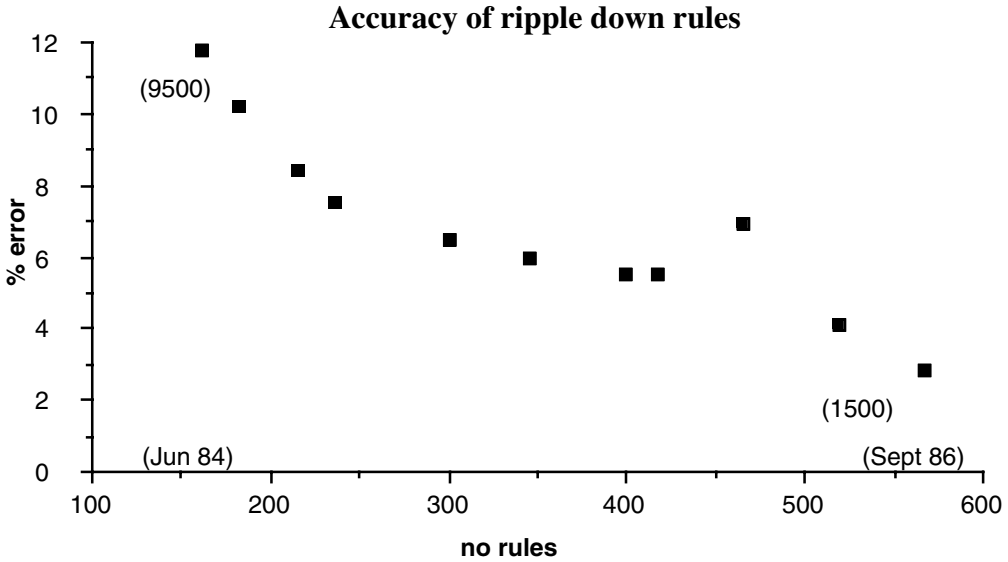

Fig 4. This figure illustrates the percentage of errors in the interpretations produced by ripple down rules when compared to the original interpretation produced by GARVAN-ES1, as a function of growth in the size of the rule base. The number of test cases reduces during the knowledge base growth from 9500 to 1500, as incorrect interpretations of test cases were used as the basis for further knowledge addition. Whether a case had been correctly interpreted or not, once it had been a even a candidate for rule changes, it was not reused to ensure the test was fully valid. The dates indicate the period the archived cases derived from.

Redundancy is obviously the major problem with this approach. Since knowledge is entered in context the same rule may end up being repeated in multiple contexts. Although redundancy has not been a major problem in the work so far it is still rather disatisfying for the knowledge engineer to have to add knowledge which he knows exists elsewhere in the knowledge base even if the addition is very easy. We are therefore investigating ways of generalising from the knowledge in this sort of

knowledge base. In our first attempt at this we have implemented a LAST_DIAGNOSIS condition which essentially uses only the information implicit in the LAST_FIRED condition specifying the diagnosis or interpretation reached. This is currently under evaluation.

Ripple down rules is just one very specific attempt at capturing knowledge in context and there are no doubt other and better strategies. But such strategies must be based on the recognition that knowledge has its primary validity in context where insight is active and that generalisation is a secondary activity. It is a mistake to throw away the specifics on the assumption that knowledge is built from generalised primitives.

Finally the reader may wonder why we did not use inductive learning techniques such as ID3 on such a large data base of cases. We have in fact used such techniques (Quinlan et al 1987), but inductive learning is not appropriate for maintenance where you are essentially dealing with the first apparent instance of an new data profile, and secondly our purpose here was to explore a new approach to knowledge acquisition and organisation.

Appendix 2- the knowledge dictionary

As well as capturing knowledge in context it is also essential that the knowledge in a knowledge base should be able to be examined in any context desired. We have noted the flexible way experts change the context in which they provide knowledge. Expert system knowledge bases per se do not allow this type of flexible access, so it is an area of major interest to develop techniques providing this type of better access to knowledge (NEXTRA etc) independent of the philosophical issues raised above.

Our approach has been to focus on the underlying representation, which may then provide a suitable basis for flexible access. One of the major issues is that the underlying representation should be normalised as far as possible. Knowledge bases have the same potential problem as large data bases. You do not want any of the knowledge to become segmented and isolated. If you have to examine or change some knowledge object, you want to be sure you have access to every occurrence and usage of that object and every relation it may have with other objects. Obviously the more the representation is normalised the more readily this can be achieved.

We have suggested (Jansen and Compton 1988a & b and 1989a) that the conventional data dictionary (Dolk et al 1987) may be expanded to include knowledge and may thus provide a suitably normalised underlying representation as well as a link to conventional data bases. The use of an extended data dictionary which includes knowledge, that is a knowledge dictionary, then opens up the possibility of applying the results of software engineering research to knowledge design and implementation.

The allowed relationship model for the knowledge dictionary is described elsewhere (particularly in Jansen and Compton 1989a). Fig 5 provides an example of this sort of representation in that rules are broken up into their component parts and stored in a relational table. This allows relational calculus to be used to manipulate the knowledge base, with all the generality and power that this implies. The example in Fig 5 is of a

semantic nets and the ripple down rules described above (Jansen and Compton 1989b). The inclusion of other representations in the dictionary is the subject of further research.

Versions of the knowledge dictionary have been implemented in both Prolog and Hypercard on the Macintosh and RDB and Rally on the MicroVax, and the GARVAN-ES1 knowledge stored in relational form into the more fully developed Macintosh dictionary. An inferencing mechanism has been developed within the relational framework, although for many applications, a run time knowledge base would be generated from the dictionary to be used with the expert system shell required.

Current work is directed towards developing the appropriate context strategies using pattern matching and table searching. For example we have previously described (Jansen and Compton 89a) a *why not* facility whereby one can query the reasons a rule did not match a certain data profile. This facility requires only simple data manipulation rather than rule parsing since the knowledge is stored as data. It is not hard to see how one may ask instead of, why not a rule, why not a specific interpretation, or certain class of interpretation? These are technical problems rather than research problems because the underlying representation supports the type of query required.

PRODUCTION RULE

RULE(42)

IF FTI is high
and T3 is high
and TSH is undetectable
and not on_t4

THEN **thyrotoxic**

ELEMENT RELATIONSHIP TABLE

owner	relationship	member
-------	-------	-------
-------	-------	-------
RULE_42	presence	FTI_high
RULE_42	presence	T3_high
RULE_42	presence	TSH_undetect
RULE_42	absence	on_t4
RULE_42	outcome	thyrotoxic
-------	-------	-------
-------	-------	-------

Fig 5. The production rule (upper) can be represented as a set of tuples in a relational table, where the rule as an object has presence and absence relationships with various facts and an outcome relationship with an interpretation. This is a simplified representation as object types etc are not indicated.

References

Ames, A Jr. "Visual perception and the rotating trapezoidal window" Psychological Monographs (1951); 65, whole no. 324.

Bachant, J. McDermott, J. "R1 revisited: four years in the trenches", The AI Magazine (Fall 1984); pp.21-32

Bateson, G (1979) *Mind and Nature: a necessary unity* Wildwood House, London

Boose, JH., and Bradshaw, JM. "Expertise transfer and complex problems: using AQUINAS as a knowledge acquisition workbench for knowledge-based systems." International journal of Man Machine studies (1987); 26 pp.3-28

Buchanan, B. "Expert systems: working systems and the research literature" Expert Systems (1986); 3(1) pp.32-51

Bylander, T. Chandrasekaran, B. "Generic tasks for knowledge-based reasoning: the 'right' level of abstraction for knowledge acquisition", Proceedings of the knowledge-based systems workshop, Banff, (1986); pp.7.0-7.13

Clancey. WJ, "Heuristic classification" Artificial Intelligence (1985); 27 pp.289-350

Compton, P. Horn, K. Quinlan, R. Lazarus, L. "Maintaining an expert system". Proceedings of the fourth Australian Conference on Applications of Expert Systems, (1988); pp.110-129

Compton, P. Jansen, R. "Knowledge in context: a strategy for expert system maintenance" Proceedings of the Australian Joint Artificial Intelligence Conference (AI'88) Adelaide (1988); pp283-297

Compton, PJ. Stuart, MC. Lazarus, L. "Error in laboratory reference limits as shown in a collaborative quality assurance program", Clin Chem, (1986); 32, pp.845-9

d'Espagnat, B. (1983) *In search of reality* Springer-Verlag, New York

Dolk, DR. Kirsck, RA II. "A relational information resource dictionary system" Communications of the ACM (1987); 30(1) pp.48-61

Dreyfus, H.L. Dreyfus, S.E. "Making a mind versus modelling the brain: artificial intelligence back at a branchpoint." Daedalus, (Winter 1988); 117(1), pp.15-43

Feigenbaum, EA. "Knowledge processing: from file servers to knolwedge servers" Proceedings of the fourth Australian Conference on Applications of Expert Systems, (1987); pp.1-10

Feigenbaum, EA. "The art of artificial intelligence: themes and case studies in knowledge engineering". Proceedings of IJCAI-5. (1977); pp.1014-1029

Griner, PF. Mayewski, RJ. Mushlin, AI. Greenland, P. " Selection and interpretation of diagnostic tests and procedures". Ann Intern Med, (1981); 944, pp.553-92

Hayes-Roth, F. Waterman, DA. Lenat, D. "An overview of expert systems", In: *Building expert systems*. eds., Hayes-Roth, F. Waterman, DA. Lenat, D. (1983); pp. 3-29

Horn, K. Compton, P.J. Lazarus, L. Quinlan, J.R. "An expert system for the interpretation of thyroid assays in a clinical laboratory", Aust Comp J, (1985); 17, pp.7-11.

Jansen, R. Compton, P. "The knowledge dictionary, a relational tool for the maintenance of expert systems", Proc. Int. Conf. on Fifth Generation Computer systems, Tokyo (1988); pp.1159-1167

Jansen, R. Compton, P. "The knowledge dictionary: an application of software engineering techniques to the design and maintenance of expert systems" Proceedings of the AAAI'88, workshop on the integration of knowledge acquisition and performance systems, (1988)

Jansen, R. Compton, P. "The knowledge dictionary: a data dictionary approach to the maintenance of expert system" Knowledge Based Systems (1989); *in press*

Jansen, R. Compton, P. "The knowledge dictionary: storing different knowledge representations", CSIRO Division of Information Technology Tech. Report TR-FD-89-02 (*also submitted to EKAW89*)

Kuhn,TS. (1969) *The structure of scientific revolutions*. University of Chicago Press, Chicago.

Lakatos, I. *Philosophical Papers* . Cambridge University Press, Cambridge@@@@@@

Lenat, DB. Feigenbaum, EA "On the thresholds of Knowledge" *P*roceedings of the fourth Australian Conference on Applications of Expert Systems, (1988); pp.31-56 (Also MCC TEchnical Report AI-126-87)

Lenat, DB. "When will machine learn?" Proc. Int. Conf. on Fifth Generation Computer systems, Tokyo (1988); pp.1239-1245

Lock Lee, LG. Teh, K. Campinini, R. "An expert operator guidance system for an iron ore sinter plant" Proceedings of the third Australian Conference on Applications of

Lonergan BJ. (1958) *Insight* Darton, Longman and Todd, London

McCarthy,J. "Circumscription - a form of nonmonotonic reasoning" Artif Intell (1980); 13:1 pp.27-39

McLuhan, M. Parker, H. (1968) *Through the vanishing point: space in poetry and painting*. Harper and Row, New York

McNamara, AR. Lock Lee, LG. Teh, KC. "Experiences in developing an intelligent operator guidance system", Proceedings of the Australian Joint Artificial Intelligence Conference (AI'88) Adelaide (1988); pp.33-40

Newell, A. Simon, H. "Computer Science as Empirical Enquiry: Symbols and Search" reprinted in *Mind Design* ed. ed. Haugeland, J. MIT Press,Cambridge, (1981)

Piaget, J. (1929) *The child's conception of the world* Routledge and Kegan Paul, London

Popper, KR. (1963) *Conjectures and refutations* Routledge and Kegan Paul Ltd., London

Quinlan, J.R. Compton, P.J. Horn,K.A. Lazarus, L. "Inductive knowledge acquisition: a case study", In: *Applications of Expert Systems*. ed., Quinlan JR, Addison Wesley, London (1987), pp.159-73.

Rappaport AT., Gaines BR. Integrating Knowledge acquisition and performance systems Proceedings of the Australian Joint Artificial Intelligence Conference (AI'88) Adelaide, (1988); pp.317-336

Rock I. (1983) *The logic of perception* MIT Press, Cambridge, Mass

Schwartz, WB. Patil, RS. Szolovitis, P. "Artificial Intelligence in medicine: where do we stand", N. Eng. J. Med (1987); 316 pp.685-8

Shaw, MLG. "Validation in a knolwedge acquisition system with multiple experts", Proc. Int. Conf. on Fifth Generation Computer systems Tokyo (1988) pp.1259-1266

Soloway, E. Bachant, J. Jensen, K. "Assessing the maintainability of XCON-in-RIME: coping with the problems of a VERY large rule-base", Proceedings of AAAI87, Seattle, (1987); pp. 824-829

Stratton, G. "Vision without inversion of the retinal image" Psychol Rev (1897); 4 pp. 341-60, 463-81 (referred to by Kaufman L. (1979) *Perception:the world transformed*Oxford University Press, New York)

Suwa, M. Scott, AC. Shortliffe, EH. "Completeness and consistency in a rule-based system" In: *Rule-based Expert Systems,* eds., Buchanan BG and Shortliffe EH. Addison Wesley, Readingg Mass, (1984); pp.159-70.

Treitel , J. (letter) Nature (1984) 308(12) p 580

Waterman, DA. (1986) *A guide to expert systems* Addison Wesley, Reading, Mass.

Winograd, T. Flores, F. (1987) *Understanding computers and cognition: a new foundation for design*. Addison Wesley, Reading, Mass.

Wittgenstein, L. (1953) *Philosopical Investigations* Blackwell, Oxford.