A DISCUSSION MODERATOR

by

G. Alan Creak and Roy Davies

(alan@cs.auckland.ac.nz)
(http://www.cs.auckland.ac.nz/~alan)

(roy.c.davies@ieee.org)
(http://www.eat.lth.se/roy.davies)

Technical Report 55

Computer Science Department,
Auckland University,
New Zealand

1991 August

CONTENTS.

A DISCUSSION MODERATOR

by

G. Alan Creak and Roy Davies.

ABSTRACT.

The suggestion that computer methods might be used to identify points of conflict between sides in an argument is investigated. It is suggested that computer assistance in moderating discussions could be a useful approach. Initially, the computer's function would be restricted to recording the arguments, as any attempt at interpreting the participants' words could lead to further conflict. A design is suggested, and a partial implementation reported.

FOREWORD

The germs of the project described in this Report are to be found in a real concern to apply the very considerable power of computing techniques developed over the past years to humane ends. There is a vast computing industry funded by governments of all persuasions, and devoted to killing people; a second empire sees computing as a way to increase commercial profits, typically by increasing efficiency and not always with humane values in mind. Computers have been used to make life easier for some people - typically the people who had quite an easy life already, and were easily able to pay for their new toys.

Comparatively little work has been done on computer applications primarily directed at humane and positive ends. Computers are beginning to find uses beyond the trivial in education and medicine, though again it is hard to avoid the impression that much of the activity concentrates on spectacular gimmickry rather than intellectual content.

We hope that our work can be seen as a serious attempt to use the methods developed in artificial intelligence under the rough heading of “knowledge engineering” to address real questions of significant public interest. We have no axe to grind, but share with many others dismay at the growing polarisation of opinion in New Zealand. We are not so naive as to think that our proposals will solve any problems; but if they give a few people pause to think, that will be better than nothing.

It is a pleasure to thank Dr Cleve Barlow of the Maori Studies Department for his encouragement. It is our regret that we have got no further with this development; it is our hope that some day we shall be able to continue.

CHAPTER 1

INTRODUCTION.

The year 1990 marked the sesquicentenary of the signing of the Treaty of Waitangi, an agreement between Maori inhabitants of New Zealand and European settlers which is of great significance in New Zealand's history. Much was made of the anniversary in New Zealand, but one of the striking features of the approach to 1990 was increased, and increasingly strident, debate about the Treaty, more notable for its rashness than for its rationality. Indeed, it seemed that the only possible conclusion to be drawn from the reported opinions on the Treaty was that there was deep disagreement over what it meant. The various communications media reported with fine impartiality the pronouncements of people with all shades of opinion and all kinds of qualifications, including none. Occasionally background articles or reviews would appear; these too, while sometimes more moderate in tone, seemed unable to agree on any significant point.

The whole spectacle was not encouraging to anyone who still retained some degree of confidence in the basic rationality of the human mind. The original documents were there to see; clearly, people's opinions about the documents' significance would depend on their backgrounds, but the differences were surely amenable to analysis by calm discussion. There was little evidence of any such approach. How could it be that well meaning people on all sides of the debate could start from the same premisses and end up at each others' throats ? Thouless[1] has made a similar point :

> *"Poetry, romantic prose, and emotional oratory are all of inestimable value, but their place is not where responsible decisions must be made. The common (□almost universal□) use of emotional words in political thinking is as much out of place as would be a chemical or statistical formula in the middle of a poem. Real democracy will come only when the solution of national and international problems is carried out by scientific methods of thought, purged of all irrelevant emotion."*

It isn't by any means a new thought. Bertrand Russell, discussing Leibniz[2] :

> *"... he hoped for an end to all disputes. 'If controversies were to arise', he says, 'there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pens in their hands, to sit down to their desks, and to say to each other (with a friend as witness, if they liked), "Let us calculate".'."*

(One wonders why they needed the friend.) Russell goes on to remark that things aren't quite as simple as that, but that a lot of them should be.

It was from the background of such cogitation that the work described in this report was born. Why should a computer be expected to introduce sanity when human efforts had manifestly failed ? There were two reasons for hoping that some good might come of the attempt. First, there was the optimistic belief that, at the bottom of the arguments, there must be logic, coupled with the undeniable fact that computers are quite good at logic. Second, there was the lack of bias of the computer, and - all being well - its software. Experience in education, and other fields, has shown that computers can be seen as impartial, non-judgmental information providers and assessors, which is perhaps just what is needed.

(In view of the present climate of hypersensitivity, under which any self-respecting socially conscious demagogue can detect social, sexual, religious, and class discrimination in any statement whatever, we add this remark on bias. We suggest that computers are things, which behave identically towards input signals no matter what their source. To that extent they are without bias. It is certain that people from different backgrounds will view the computer in different lights; but different backgrounds are equal in that the computer is equally alien to all. We are not going to insist that the participants in the debates we discuss should sit down at computer terminals; it would be much more efficient to have material entered by people who can type quickly. Direct speech input would be even better, but that isn't practicable at the moment. The bias we mention is to do with logic. The computer will check all logic equally rigorously; it will not accept skimpy arguments from one side on the grounds that they are "obviously" correct, while being fiercely critical of those from some other quarter. Of course, if you believe - as some claim to believe - that logic itself is somehow discriminatory, then we part company. But that has nothing to do with computers; if you reject logic, then we can't talk to you as people either.)

What, then, should the computer assess, and what information should it provide ? Early ideas, recorded in two unpublished notes[3,4], were directed at the possibility of identifying the roots of the disagreements by analysing the arguments through dialogue between the participants. Eventually it was hoped that this would lead to the identification of systematically different viewpoints held by the two parties, which, once brought to light, could be discussed. The approach was inspired by Bartlett's work on schemata, described in this quotation[5] :

> *"The first case is that of schema theory, which originated within psychology in the work of the British psychologist, Sir Fredrick Bartlett, although the roots go back at least as far as Kant (1787). In the 1930s, Bartlett examined the ways in which people distort and reconstruct the memory of some event or story which they have previously heard or read. In one of his most famous studies he used a story based on a North American Indian legend 'War of the ghosts'. He gave this story to people to read and then tested their recall of it after various intervals of time. Bartlett was concerned with the systematic memory errors which non-Indians made in recalling the story (he had deliberately chosen a story which did not fit with the cultural conceptions of the people in his experiment). His subjects 'forgot' aspects of the legend which were incompatible with their knowledge. To account for his findings, Bartlett proposed that when people read a story they construct an abstract representation, or schema, of the story's general theme. This representation, he proposed, is affected by the reader's personal system of beliefs and emotions."*

This seemed to fit well with the Treaty problems. It suggests that people were not, in fact, discussing the Treaty at all : they were discussing what they thought it meant, which could differ from what it actually said because of the modifications induced by each person's "personal system of beliefs and emotions". If that were the case, then - we argued - by analysing disagreements in interpretation of the same document we should be able to identify at least the differences between different people's fundamental assumptions. That wouldn't solve any problems; but knowing where the problems originated could help reasonable people to come to an accommodation of some sort.

(It seems, incidentally, that "schema" is nowadays something of a rude word in psychological circles, doubtless because it's been used indiscriminately in ill-defined contexts. It still seems to be acceptable in artificial intelligence, though, and it turns up in a lot of books, so we'll stick with it.)

Of course, it is far from uncommon for people to hold different views on topics, and for these differences to remain unstated. Ramirez[6], in quite a different context (□work towards developing an expert system involved in controlling an aluminium plant□, makes the interesting observation "The experts selected different rules for the same problem. What initially was interpreted as inconsistency, however, was produced by differences in the contexts and their interpretation". Here again we see the importance of viewpoint on conclusions drawn even by experts. It is even common for one person's opinions to be inconsistent. From R.H. Thouless[7]□:

> *"We can all have many inconsistent opinions whose inconsistency we are not able to recognise until someone else shows it to us. I was once doing an experiment on a group of people to find the degree of self-consistency of their opinions. The amount of inconsistency revealed in their answers was surprising. For example, a large number both asserted that every statement in the Bible was literally true and also denied that Jonah emerged alive after having been swallowed by a great fish. It was not that they did not know that there was a statement in the Bible that Jonah was swallowed by a great fish and afterwards came out alive; it was merely that these were two opinions which they had formed separately without relating them to one another."*

WHAT IT'S ALL ABOUT.

The original object of the exercise, then, was to determine by analysis of people's arguments the differences between their fundamental opinions about some topic - specifically, the Treaty of Waitangi, though there is no reason why our methods shouldn't apply to any other area. That means that we would have to get some arguments, and analyse them. We could try to do that in at least two, not necessarily mutually exclusive, ways.

First, we can adopt an *analytical* approach : start with the argument and work backwards. Given enough material, we might hope to be able to find out how people justify their conclusions, and from these justifications derive individual schemata incorporating personal axioms not necessarily accepted by other people.

Second, we can adopt a *synthetic* approach : try to model a person arguing. In this case, we aim to produce a system which will start from an appropriate set of axioms, and work through the logic to reach the same conclusions as the person in question. A part of the set of axioms will be a supposed schema for the person we are trying to model; we accept schemata when the answers generated by the programme agree with the people's.

In either case, we would be unlikely to be able to work from the original texts of the data. In order to recognise the fine shades of meaning which we would need, we would require natural language understanding techniques of a subtlety not yet achievable in practice. We would therefore perforce expect any arguments considered by the system to be couched precisely in some formal notation, such as propositional or predicate logic.

As we shall explain, this original intention, set out in detail in Chapter 2, did not survive for very long. It became clear on considering our first experiment (Chapter 3) that our preference for formal notation could not be satisfied without introducing complications. Specifically, we were not convinced that people's normal approach to discussing issues could be formalised precisely, and we were afraid that any attempt to do so would distort, or at least appear to distort, the debate. Our analysis could then be disputed by anyone who cared to claim that our formalised record did not in fact capture the real meaning of what had been said. We therefore redirected our efforts to recording what was actually said in the debates, and the links which the participants themselves saw between the statements. Chapter 4 describes an experiment performed under this regime, and Chapter 5 reviews all the work reported.

A PLATONIC POSTSCRIPT.

In Plato's *Republic* he has Socrates say, after Thrasymachus has expounded his own view at considerable and tortuous length[8] :

> *"We might answer Thrasymachus' case in a set speech of our own, drawing up a corresponding list of the advantages of justice; he would then have the right to reply, and we should make our final rejoinder; but after that we should have to count up and measure the advantages on each list, and we should need a jury to decide between us. Whereas, if we go on as before, each securing the agreement of the other side, we can combine the functions of advocate and judge."*

The pattern of debate advocated by Socrates in this passage is stepwise : a contributor puts forward a point, supports it from earlier conclusions of the argument or otherwise, and proceeds only when all contributors accept the point made. We would like to regard the work described in this report as an attempt to provide administrative help to someone moderating such a debate. Socrates, of course, did a lot more than administer : with the advantage of a lot of help from Plato, he controlled the debate, and took part in it himself. We don't (yet) aim to emulate his performance, and in this introduction we use his example to delineate just what we have attempted, and to outline how it could fit into a wider, more fully Socratic, system yet to be developed.

To illustrate the point, we may regard the functions of Socrates as discharged by two components, Orts and Scae. They are interleaved and interdependent (which is just as well for Socrates) in Plato's presentation, but we have separated them. Orts is a programme which records the developing structure of a debate; it is this programme which we have attempted to construct, but not yet managed to complete. It is only half of Socrates, and a slightly inferior half at that, because, while it encourages and administers a Socratic form of dialogue, it does not contribute anything of its own to the debate. Indeed, as we shall explain later, we have been at pains to ensure that it *cannot* contribute anything, so that there is no chance of the argument being changed by the computer. We may, some day, feel competent to write Scae, the more interesting part, which understands and joins in the debate; but at present we expect that day to be far off. We return to this matter in Chapter 5.

CHAPTER 2

FIRST FORMULATION OF THE PROBLEM.

Our object was to construct a reasoning system which included an effective representation of the Treaty of Waitangi and of schemata corresponding to the attitudes of some individuals. Even accepting that it was impracticable to attempt natural language understanding at the level of subtlety which would be needed for such a sensitive issue, we still saw the function of the computer as primarily a matter of the logic of the arguments.We would build a logical structure as the debate proceeded, adding new components to represent each new proposition and step in the argument. We hoped that the computer would help by keeping the discussion honest - by making sure that the same statement could not be used in inconsistent ways, perhaps by searching for contradictions in its structure, and drawing them to the attention of the debaters, or perhaps simply by making each step in the argument plain and open to scrutiny by all concerned. The result should be to ensure that any significant shade of meaning would be explicitly recognised, so that puzzling differences in interpretation apparently deriving from the same premisses could be resolved. All this is a matter of straightforward logic, and we did not expect that it would be difficult to implement.

The tricky bits are those connected with the Treaty itself, and with identifying and generating the schemata. The "shades of meaning" and "differences in interpretation" mentioned in the previous paragraph presumably had something to do with the schemata, but it wasn't clear how to derive the one from the other. Neither was it clear how we should encode the Treaty; should we try to represent it in terms of formal logic, or should we regard it as something to be debated, but not directly represented in our system ? It seemed likely that an effective system would need a formal representation of the Treaty, particularly if we hoped to adopt the synthetic approach mentioned in Chapter 1.

The synthetic approach seemed attractive, as it offered a solution to the problem of defining the schemata. To make this work, the system must also provide facilities to give interpretations of the Treaty as it applies to matters of concern in its range, and interpreted through a selected schema. We could then adjust the schema, by means to be devised, to match the person's opinions. How would we be able to tell whether the system were performing this job satisfactorily□

CONCEPTUAL DEPENDENCY.

Despite the title of W. Lehnert's book *The process of question answering*[9], that question is in some ways its central theme. Lehnert discusses ways of representing the information conveyed by short stories, and of answering questions about them. Her questions and answers can be seen as means for testing the effectiveness of the representation. We face the same need to test a representation; we have little, if any, choice but to adopt the same solution. In other words, we can only test our system's ability to interpret the Treaty and the schema by requiring it to do so□by offering it test cases, and inspecting the results it provides.

We do not need to go to the lengths implied by Lehnert in her book, as our aims, leaving aside for the moment any complications which may result from the subtleties of the subject material with which we propose to work, are less ambitious than hers. The system she describes has three clearly distinguished components□the English language input analysis part, the question-answering component itself, and the English language generator for the answers. As we intend only to handle the reasoning connected with the arguments, and not the linguistic features, we do not need the first and third components. This means, as we have seen in Chapter 1, that we must be prepared to express our questions and information in some formal style acceptable to the interpreting programme; but for our purposes this is perhaps even an advantage, because of its precision. We are trying to detect shades of meaning, small semantic differences between statements and points of view, and it would be very hard to have to extract these from natural language before we even begin to analyse them. It is significant that far the greater part of Lehnert's book is devoted to the niceties of understanding English, and comparatively little to the process of deciding on what sort of answer to give once the question has been identified. We are therefore interested in Lehnert's QUALM[10], and less so in SAM and PAM, the natural language components of her system. Even within QUALM, much attention is given to evaluating the conversational context in order to decide just what aspects of the knowledge of the situation should be reported as the answer[11]; we don't need that either.

> It would nevertheless be interesting to study the natural language interfaces to the question-answering system, particularly in view of our concern for two languages. An advantage in principle would be the computer's comparative impartiality : even if its programmes are biased, it does the same things to both sides of the debate. Lehnert's own system has been operated to some degree with several different languages[10]□ without any change to QUALM, which is gratifying□ so the neutrality of the technique is established.

Lehnert's model is strongly based on the *conceptual dependency* techniques introduced and developed by Schank[12]. The important feature of this approach is that it handles everything using a common set of primitive notions, particularly actions[13].

Schank emphasises the importance of an *interlingua*[14] as a medium in which thoughts can be expressed unambiguously and uniquely, independently of normal language, so that surface differences between alternative ways of framing the same thought in English (□or, in principle, in any other language□) will disappear. Much of his concern, though, springs from his aim of coping with normal speech. This leads him to be wary of accepting words as units of expression□ words have synonyms, or□ worse□ near synonyms; they are rarely precisely defined; and their meanings may be complex. He therefore prefers to define his own units of representation, which are his semantic primitives. We share his need for precise specification, but not the constraints imposed by natural language. In Schank's terms, we require that everything be encoded in interlingua anyway. As well as that, we need not aim to cover all possible subject material and topics of discussion. We can therefore restrict our definitions to things we find it necessary to define, and construct these definitions carefully to suit our requirements. We can even retain the convenience of words, provided that we guarantee that they will always be used with precisely the same meaning.

Schank's work is focused on actions, his primitive components of verbs. Stripped of its (□sometimes confusing□) diagrammatic representation, his approach is essentially a case grammar technique[15]. He is less careful in his treatment of nouns, where his descriptions are sometimes less than convincing. For example, he states that[16] "'Doctor' is mapped conceptually into a PP (□human□) plus other information describing the particular types of human that a doctor is". But why□ Why should not "Doctor" be mapped into "animal", or "thing", or "professional practitioner"□ or, indeed, into "doctor"□ Some light is thrown on the matter by Rieger[17], who speaks of the mapping as an association "between a token and its 'least biased' classification"; we are to choose the class which is of "*general* utility or interest". In other words, the decision is arbitrary, and is governed by the context in which we wish to use the knowledge. Lehnert[18] extends the idea to conceptual primitives for physical objects, in effect classifying objects into the categories setting, gestalt, relational, source, consumer, separator, and connector. Each primitive is a set of potential properties, typically described by scripts or frames. For example, every setting may have scripts describing what happens there, and frames describing typical contexts, every source may have scripts describing how it is activated, and frames describing the nature of the substance provided, every separator or connector may include information on what is separated or connected, and so on. An object is described by regarding it as acting as one or more of the primitives. In this way, properties associated with an object which are directly concerned with its function are expressed in a clearly identifiable and accessible way.

How does all this relate to our proposed work on the Waitangi Treaty□ Mainly in its prescription of the way in which entities interact. We are concerned more with interacting objects than with activities (□even if we wish to investigate the position of an activity under the Treaty, we are really concerned to determine whether the effect of the activity is to produce an illegal state of affairs□), and the idea of classifying objects by primitives, which list their important properties in a quite general way, seems potentially very useful. Of course, there is no reason to suppose that Lehnert's primitives will be appropriate to our needs, but the principle they embody may be very valuable. The schemata fit in as scripts; they describe expectations about the ways in which things are related. A novelty in our case is the provision of separate scripts for different people, but this requires no bending of the principles of conceptual dependency theory.

A feature which will be central to our work[19] is the use of consistency checks. That's about how to work out that you can't answer a question, and to take sensible action. QUALM didn't do that; but it's exactly what we need to initiate new searches for bits of structure we have to provide. We shall see that in many cases the opportunity for introducing personal values into an argument comes in through loosely framed logic.

Of the material on conceptual dependency, the most interesting for our purposes (□nless we decide to tackle the natural language problem□) is that of Rieger[17]. He describes the implementation of the inference mechanism of a conceptual dependency system. The details of Rieger's programme are determined by the context, but the important structure is guided by the aim expressed by Rieger as "to determine points at which one pattern joins with another pattern". That is our aim too. Given a pattern representing a question, we have to identify within the body of material at our disposal points at which the pattern of the question matches the pattern of stored knowledge. Our machinery is likely to be rather different from his, though, as his system is driven by a "central reflex response", with directions to search governed by commonsense questions about the new material. For example, given a sentence like "X went out with a screwdriver", it would perhaps explore reasons for going out and applications of screwdrivers. We do not need this sort of search; we can specify quite precisely the path of argument to be followed in any line of reasoning; what we require is the assurance that the path can be repeated if we change the underlying knowledge base, and a clear indication of why it can't if that turns out to be so.

That suggests that Rieger's data structures are more likely to be useful than are his control mechanisms. The major data structures[20] for objects are property lists, with each property value being a complete statement□so the value of a X's *employer* property could be the statement "X is employed by Y". These assertions are in principle independent of the property lists, so the same assertion could also be used as an *employee* value for Y, thereby linking the whole structure together. Inheritance is provided through a single *is-a* hierarchy.

KL-ONE AND KRYPTON.

An alternative, though not entirely dissimilar, approach is seen in the development of the KL-ONE[21] system for knowledge representation, and particularly in its descendant Krypton[22]. These languages are essentially frame-based with inheritance. The special feature of Krypton is its careful distinction between two sorts of statement□the *terminological* and the *assertive*. Roughly speaking, a terminological statement serves to define words in terms of other words, and records something which is believed to be essentially true (□A wristwatch is a sort of clock"□), while an assertive statement records a fact which happens to be true (□There's a clock on the Town Hall"□). A careful treatment of the semantics of the system leads to an implementation in which the two sorts of information can be effectively used in conjunction.

LEGOL.

Yet another approach to the use of computers in handling arguments may be seen in the various versions of Legol[23], which is described as "a language for writing rules such as those which might appear in legislation or system specifications, in such a way that they can be interpreted automatically and tested to discover whether they will have the desired effect". This is doing just the sort of thing we want so far as interpretation goes; but it is formulated as a quite traditional programming language, and appears to give little scope for incorporating anything corresponding to our schemata. A later proposal[24] seems to be more flexible, and explicitly addresses the need for greater complexity□specifically, to consider the effect of several interrelated regulations in parallel.

One idea from Legol may be significant in our work. A single object in Legol may be represented by several different entities, each corresponding to the object in a particular period of time. As we may wish to distinguish between circumstances before and after the date of the treaty, time is of importance to us too.

A paper by Sergot and others[25] describes an outgrowth of the Legol project which has several distinctive features. Its most distinctive point is its implementation in Prolog, a declarative language which is designed for logic programming; Legol, in contrast, is firmly rooted in the tradition of procedural languages and database methods. (□It is interesting that the authors' stated motivation was quite like ours□ "We hoped that formalization of the various definitions might illuminate some of the issues causing the controversy".□) In both cases, the legal framework is seen as a tree structure, but whereas in Legol the rules are applied to an individual in a bottom-up way as guided by the code, in the Prolog system the rules are (we suppose) applied top-down as they are needed. At first sight, the greater flexibility of Prolog could fit it better for our needs; but Sergot and colleagues remark on difficulties attached to Prolog's interpretation of negation.

The Prolog group also report on their development of the Prolog representation by a process of trial and error, and remark that this may be an inevitable consequence of the nature of legislation itself. The process they describe is very like that we propose to use. They exhibit a session with their implementation, showing how the programme asks questions to acquire information it needs, and also how it answers "why" and "how" questions.

BUT IS IT POSSIBLE IN PRINCIPLE ?

The Prolog paper led to an interesting sequel. Leith[26] criticised the paper by Sergot and others (□which he confusingly calls "Cory et al", having reversed the list of authors' names in his cited reference□) on legal grounds. (□He also speaks of "formalisation" and "normalisation" in such a way as to blur any distinction between the two□ we shall assume he meant "formalisation" throughout.□) His point is that the same law may be interpreted differently by different legal authorities, even in circumstances which appear to be quite similar, and that therefore logic programming is an inappropriate approach. Accepting that his premiss is valid, it is far from clear that he understands the nature of logic programming; certainly, to base his arguments essentially on the Prolog paper was unwise, as this is self-confessedly a simple example, "free, for the most part, of many complicating factors that make the problem of simulating legal reasoning so much more difficult". Kowalski and Sergot[27] have answered Leith's criticism. So far as our own proposed activities are concerned, we find that Leith's point is at the centre of our proposal ! It is because we are concerned at the differences between interpretations of legal material that we have embarked on this study; and it seems that the examples which Leith quotes could very well benefit from an approach such as ours, where we seek to identify the roots of the different interpretations and exhibit them for evaluation.

Berman and Hafner[28] look at the benefits which legal practice might hope to gain from artificial intelligence. They also mention the inconsistencies and indeterminacies which are common in legal decisions, but, like Sergot and colleagues, direct their attentions to coping with the law as it stands. They believe that expert systems of two sorts could be useful□ *predictive* expert systems, which could construct informed guesses as to the outcome of litigation, and *normative* systems, which could match the circumstances of a particular incident against records of related incidents, and help in a decision on sentencing. They do not consider the possibility of using artificial intelligence techniques to analyse the effect of law, and the reasons for the inconsistencies they observe; that is our concern.

We may also question the possibility of expressing our arguments in a formal way. Whether we follow Schank and speak of an interlingua, or just assume that what we want to say can straightforwardly be encoded in formal logical terms, we are in some sense limiting the vocabulary which we can use. The dangers inherent in such formalisation have been most dramatically explored by George Orwell[29]. For example[30] :

> *... in Newspeak the expression of unorthodox opinions ... was well-nigh impossible. ... It would have been possible, for example, to say* Big Brother is ungood. *But this statement ... could not have been sustained by reasoned argument, because the necessary words were not available.*

Perhaps Schank's interlingua, and the constraints of formal logic, are not as restrictive as Newspeak - but they are by no means the currency of our normal thought, and to that extent may introduce distortion. Whether or not that distortion is serious we can only determine by experiment.

THIS IS WHERE THE STORY REALLY STARTS.

This fairly selective survey of the literature suggests that, while much attention has been given to understanding natural language, representing the meanings of sentences, and handling legal documents and arguments, none is directly related to our problem - indeed, it is only mentioned as a complicating obstacle to progress, so any solution would be of value. We have found no material at all relating to parallel debates in different languages.

Our suspicion that interpreting natural language would be too crude a tool for our delicate requirements seems to be supported. It does not seem likely that we would be able to translate the material of debates on the Treaty of Waitangi into terms of Schank's primitives without distortion - and, in any case, the resulting representation is far too hard to interpret to tell whether or not it is accurate. Requiring the participants to express their arguments in terms of a formal representation seems much more promising, and is certainly much easier to check.

Does that amount to a polite way of saying that we find our reading to have been useless ? Not really - although it's true that we shall not make use of any of the specific techniques we found in constructing our system. The best solution is still to use a representation based on natural language, and if we can eventually augment our system by some language understanding techniques, however primitive, we might be able to make it better. What we have concluded is that at present these techniques, whether of actual language processing or of representing meaning, are unsuitable for our main purposes; but that leaves our system restricted to manipulating uninterpreted character strings, and we would like to do better than that in the long run.

SO JUST WHAT SHALL WE DO ?

We base our procedure on the assumption that the logical implication

$$\text{Treaty \& Knowledge \& Schema} \Rightarrow \text{Opinions}$$

is correct, and try to identify the schema by considering the other three terms in the equation. It is unlikely that anyone will actually present us with an argument expressed in that form, so we will need a procedure with which we believe we will be able to identify the various terms.

Significance of the Treaty.

The appearance of the Treaty in the implication is obviously necessary, given our intended field of application, but there is more to it than that. Our intention is that each step of the argument should be supported by what has gone before and, in principle, accepted by all parties before the next step is addressed. What does that say about the initial state of the argument ? It implies that there must be some body of agreement before the argument starts on which we can construct the required justifications. (This agreement may be formal rather than actual - Plato[8] frequently has Socrates rely on argument by reductio ad absurdum - but the principle is unchanged.) In the implication above, the schemata are by definition individual, and knowledge, while thought to be factual, may be wrong; the Treaty is the only common ground we can guarantee between the participants.

That being so, we might guess that even if, in some discussion, we had no Treaty, we would have to find some common ground from which to begin. The importance of this preliminary step was brought home to us as we carried out the experiment described in Chapter 4; we didn't define any common ground, and in consequence got nowhere.

Finding the schemata.

Given an argument step, how do we identify any schema component ? About the only thing we know about it is that it will, in the nature of such schemata, probably be unstated. That means that our task is to explore the unstated assumptions behind the argument steps. For example, a statement to the effect

If he stole the car then he will be locked up

implies the assumption that people who steal cars will be locked up. More subtly,

I am out of work;
people out of work should receive unemployment benefit;
so I should receive unemployment benefit

is a valid argument, but all manner of assumptions may lie behind the major premiss (the middle line). Even in this example, the questionable step is fairly obvious : the minor premiss ("I am out of work") is presumably readily verifiable, and the appearance of a word like "should" is a clear pointer to a set of criteria underlying the judgment. Unless these have already been established in the argument, they should (!) be made explicit. Yet another potential source of trouble is the authoritative statement presented as true; as an example, for some years one of us was led astray by an authoritative statement from a trusted informant to the effect that

No Maori was ever punished for speaking Maori at school.

The informant was not trying to mislead - he clearly believed what he said to be true - but evidence conflicting with his statement comes from Maori people who were themselves punished for speaking Maori at school, and one can reasonably judge their personal experience to outweigh the informant's belief. (Perhaps, indeed, he could better have said

No Maori should ever have been punished for speaking Maori at school

which puts his assertion back into the "should" category.)

Faced with such a battery of various and subtle means of injecting personal views into an argument, how can our system identify them ? The only safe way is to query every statement which is not properly supported by argument, or accepted as an axiom. In practice, that would be silly. All arguments about commonplace matters rely on a vast collection of unstated assumptions which are logically indistinguishable from those which go to constitute the schemata. The only difference is that they are not in dispute; we call the collection *common sense*. There is no immediate possibility of encoding such a mass of information (though Lenat[31] and colleagues, in their Cyc project, are attempting to do just that), so we have to find ways of avoiding the confusion.

The only devices at our disposal which come close to being able to exercise the subtlety we require are people, so we have little choice. We have more freedom in deciding which people to use. The alternatives are to rely on the participants in the debate, or to introduce a monitor with the task of distinguishing opinion from accepted or verified fact. Without engaging in a deep study of the consequences, we remarked that to involve a third party in the debate was to invite yet another potential source of bias, and therefore chose to leave the matter to the participants, and to attain our ends by relying on critical appraisal of the arguments presented.

Notice that this procedure will not waste time on expanding any assertions which command universal acceptance. If everyone is happy that “people out of work should receive unemployment benefit”, then we shall never probe into its justification. This is sensible enough - but it does leave open the possibility that not everybody means quite the same thing by it. For example, some might literally mean “everybody”, from babes in arms upwards. If we accept the procedure, then, we must also accept that we might have to reconsider statements or argument steps which happened a long time ago in the argument, and we must have ways to deal with any changes in definitions which ensue.

Roughly, then, our chosen method is to encourage the parties to the debate to criticise each other’s arguments, and for each to require the other to justify any doubtful assertion either in terms of further assertions or by accepting them as axioms. When no further analysis is possible, then - perhaps - we have our schemata.

CHAPTER 3

AN IMAGINARY EXPERIMENT.

As an illustration of what might be expected of such a programme, we devised what could be considered (□by anyone already following a Socratic train of thought□) as a Platonic ideal example of how it should work. The intention was to determine what sort of constructs we might find in an argument, and therefore what facilities our system would need - particularly in the matter of data structures - to cope with them adequately.

The example is presented in the form of a dialogue. While there is nothing in the design to determine the order of the participants' contributions, alternation is convenient, and more likely to concentrate attention on a single line of argument. The alternative is to accept lengthy arguments from each of the participants; but, as we saw in Chapter 1, the disadvantages to this procedure were recognised by Plato.

We characterise the *state* of the argument by listing the propositions accepted and not accepted by the participants. For the sake of convenience, we separate out the propositions on which they both agree; then what remains is presumably in some way related to their different schemata, and it is this body of disagreement on which the debate should focus. We do not include in the state any idea of whose turn it is, as we saw in the previous paragraph that the order of contribution was not an essential part of the argument.

The experiment begins with two items already defined : an agreed basis for argument, in this case the Treaty of Waitangi; and a disputed statement. It is necessary to begin with a controversial proposition (X), because the disagreement is the driving force of the algorithm. If we begin with agreement, there's no obvious way to decide what to do next. (This becomes very clear in Chapter 4, where we - inadvertently - try the experiment.)

The discussion is supposed to proceed between two people, imaginatively called A and B, who are called on to speak by Z, described as an administrator. Z takes no active part in the debate, but controls the order of events by directing the next speaker to comment on some particular feature of the current state of the system. The administrator is a convenience to point the way of the discussion; whether such a functionary should be part of a real debate system is questionable, but there should probably be some agent charged with the responsibility of ensuring that the discussion proceeds along a coherent path without becoming hopelessly bogged down in side issues. It may be that Z's function can be automated; on the whole, this seems to be an ambitious aim, and we later speak of a moderator.

THE EXAMPLE.

Here is a sketch of how we expect the consultation process to be conducted. We present it as an interview for illustrative purposes only; in practice, I'm sure it would have to take a lot longer, if only to allow the participants time for careful consideration of their contributions.

Dramatis personæ :

A, B : two people.
W : a treaty, accepted by both A and B.
X : a proposition claimed to depend on W, accepted by A but denied by B.
Z : an administrator.

INITIAL SYSTEM STATE :

Not in dispute :	W.
Accepted by A :	X.
Not accepted by B :	X.

NOTE : B does not necessarily accept X.

Z : A, please explain why you accept the proposition X.

NOTE : It would be more satisfying to make the process symmetrical by asking both A and B to justify their positions; but, while someone holding an opinion may reasonably be asked to justify it, and to respond by deriving it from some agreed axioms, you can't very well ask someone not to derive a result from axioms. There is a special case in which B claims to be able to justify the negation of a proposition asserted by A; in this case, B can take the initiative. An example follows later.

NOTE : Is it reasonable even to expect that A will be able to justify X ? It is, after all, quite common for people to hold deep convictions which they cannot justify by any rational means. We suggest that the problem is so constrained that A should be able to provide a justification - because ex hypothesi X is "a proposition which depends on W", and the nature of the dependence is the justification we seek.

A : W □ X1; X1 □ X2; X2 □ X.

SYSTEM STATE :

Not in dispute :	W.
Accepted by A :	X, W □ X1, X1 □ X2, X2 □ X.
Not accepted by B :	X.

Z : Thank you. B, please comment on A's chain of reasoning.

NOTE : That is a reasonable request. If B accepts W but denies X, then B must disagree with at least one of the steps in A's argument.

B : I do not accept X1 □ X2; I accept the other steps …

SYSTEM STATE :

Not in dispute :	W, W □ X1, X1, X2 □ X.
Accepted by A :	X, X1 □ X2, X2.
Not accepted by B :	X, X1 □ X2, X2.

NOTE : We are back where we started, with a proposition accepted by A but not accepted by B. If nothing else happens, we can use the same procedure recursively. But -

B : … but I can demonstrate that X2 is false.

NOTE : B has taken the initiative. If this never happens, the onus remains on A to justify the original argument in as much detail as possible. The terminating case of the recursion is illustrated later.

Z : Please present your demonstration.

B : W □ Xa; Xa □ ¬X2.

SYSTEM STATE : getting too complicated to display, but you see the idea. Other things will be necessary : they remain to be identified, but it seems likely that each sequence of deductions should be saved against possible later reassessment. For example, if B successfully convinces A of the invalidity of X□□ X2, then A will (□presumably) wish to review all arguments including that step, and all conclusions depending on it.

Z : Thank you. A, please comment on B's chain of reasoning.

NOTE : The tables are turned; the recursive pattern continues, but with the participants interchanged. This can go on for ever, but I shall take this opportunity of illustrating the recursion termination step.

A : I agree that W □ Xa, but not that Xa □ ¬X2.

Z : Thank you. B, please justify Xa □ ¬X2.

B : I cannot. For me, that is an axiom.

NOTE : We have to accept that. We all have axioms which we cannot justify rationally, as without such axioms we would have no basis for any arguments at all. Axioms need not be very deep : B could have answered "I believe Xa □ ¬X2 because I saw it happen" (appeal to experiment) or "... because C, whom I respect, told me so" (appeal to authority). In any case, if we get to an axiom, we have in some sense succeeded. It is presumably (as A has rejected it) a part of B's schema. I am not sure how to proceed; we can ask A to comment, but unless A can in some way disprove B's axiom there's nowhere to go. Observe incidentally that A must already believe that B's axiom is false, because A believes Xa and X2; any serious attack on B's position must be in terms of steps accepted by B.

SOME REMARKS STIMULATED BY THE EXPERIMENT.

- This may take some time : it's unavoidable. Trying to identify honest differences of opinion rather than score political points *does* take a long time.

- Despite our earlier remarks, the computer has not in fact used any logic at all. Do we really want it to ? We may approach an answer to that question by asking another : what can a deeper computer treatment offer ? It can check for incidental contradictions, both within one participants' beliefs and between participants; it can check for circular arguments, which would otherwise confuse the recursion; it can act as an information retrieval system. It cannot check individual steps of argument which depend on general knowledge in any simple way : how would you check "I want healthy teeth □ Fluoride should be added to the water supply" ?

- The belief structure developed will be complex, particularly when several participants are involved and several questions are under consideration. Its administration is complicated by the need to allow people to revise their opinions. (It may never happen, but just in case ...)

CONCLUSIONS.

At the end of the experiment, we have a certain number of propositions acceptable to all participants, and a set of others acceptable to some but not to all. By matching people’s sets of beliefs, we may be able to identify individual or group schemata - or we may just find that we have a set of information without much structure.

A major comment must be that people don’t, in fact, argue like that. They do not manage to preserve the Olympian detachment needed carefully to formulate contributions as syllogisms. As we saw in Chapter 2, real arguments are characterised by incompletely specified premisses, by steps which rely on evidence which hasn’t yet been introduced (or on none). Putting that another way, it is inadequate to represent an argument step as X1 □ X2; it is much more likely to require some extraneous input, as in X1 & E1 □ X2. There are now many more targets for attack in the next step, as all components of E1 as well as the implication itself are new. The potential branching factor of the discussion is therefore much larger than might appear from the example.

Perhaps this observation may lead us to qualify the opinion expressed earlier that formal expression is a necessity if computer implementation is to be useful. Logic is not the only rôle for a computer; there is also likely to be a massive purely clerical task, in keeping track of who said what and when, and how the various assertions link together in the participants’ several arguments. Indeed, we saw that it is the clerical task alone which the “computer” in our experiment has performed, even though the logical task could hardly have been expressed in simpler terms. As we remarked in the previous paragraph, people do not in practice say “W□□ □Xa”; they say “Well, obviously, if you drop the glass the drink will spill”. Unless we go back to our natural language processing, which we decided in Chapter 2 to be too unreliable, somebody has to translate the English into the logic, but it seems wise to preserve the English in case the translation is called into question. Perhaps, then, we should provide for both the original text and a more formal expression if necessary; then the computer implementation can work on the formal structure, but still be correctable if need be.

To spell out the details, observe that to derive such a formal representation (say, *f*(text), etc.) is itself to propose a step of argument□

text □ *f*(text)

This step should be open to comment and attack in just the same way as the others. And that may be enlightening ! - but it may equally lead the debate astray into a wilderness of arid definitions.All we need is someone else to insist that *g*(X) is a better representation than *f*(X), and the debate shifts from the Treaty of Waitangi to the semantic properties of different formalisms; and, so long as we place any serious reliance on the logical evaluation, there is good reason to insist on getting the best representation we can, because it will affect the conclusion. It begins to look as though our hopes for a purely logical treatment were unwarrantedly sanguine.

We, rather unwillingly, conclude that we should regard the computer logic, if at all, as a source of suggestions as to possible logical faults or loose ends which should be cleared up, but rely on the original text and on our own understanding of it for the real argument. We henceforth concentrate on the clerical and administrative tasks, where we believe the computer can offer valuable assistance. Help with the logic too would, obviously, be welcome, if only to identify those parts of the discussion still without justification, and to seek for conflicts in the undergrowth, but to automate the clerical task alone would be a significant contribution.

CHAPTER 4

A REAL EXPERIMENT.

Encouraged by the admirably orderly argument in our abstract explorations, we decided to take another step towards realism. The basic model followed in the previous chapter seemed likely to work reasonably well, modified as suggested by the conclusions presented at the end of the chapter. We decided, therefore, to embark on a real experiment, with a real discussion between people, and recording the real text of what was said.

Well, perhaps not *quite* a real experiment : we still didn't have an implementation. Nevertheless, it was an experiment of a sort, directed at exploring the consequences of our new, less ambitious, aims. We expected that the main difference from the experiment of Chapter 3 would be the appearance of real statements rather than symbolic logic in the arguments put forward by the contributors. It was also more realistic in that, rather than synthesising a discussion in the abstract, we collected together a group of people (actually, the three of us) and worked through a pencil-and-paper run of the scheme as outlined. (It would be more precise to say that we *intended* to work through such a run, but like so many other good intentions, it didn't quite happen. Perhaps that's realistic too. Details follow.) We departed from realism in that we expected to stop - probably rather frequently - for discussions on ways and means, observed defects, possible improvements, and anything else that might seem useful.

We needed a topic for discussion. We required a proposition on which there was some disagreement, but not one which was very important. It would be something with enough meat on it to explore the way the system works, but we didn't want our participants to feel they must take a long time working out answers. In the absence of expert advice, we thought it better to base our "debate" on some topic other than the Treaty of Waitangi. Instead, we chose as our subject "Should the health service be privatised ?", largely because it was a question which we all understood, and on which we could identify two clearly opposed views. This change of topic turned out to have consequences which we should have foreseen, but didn't.

In our planning, we classified the people concerned as *discussers* and *technicians*. The technicians would simulate the computer programme, which would be used by the discussers as they discussed. We constituted ourselves as a group of discussers. We didn't allocate specific rôles to individuals; instead, we all took all parts, and fabricated questions and answers with some regard for realism, but also to highlight questions which we wanted to answer.

We didn't bother to appoint a group of technicians. We thought that, with the three of us around, it would be easy to discharge all required functions. In consequence, nobody executed the programme, and we got lost. Read on.

After each stage (question, answer, interjection, whatever), we paused for discussion, and considered what part the computer system could play in the debate so far. We had intended as far as possible to follow the scheme of Chapter 3. In the event, we didn't, which was illuminating in itself. The account which follows is a tidied-up version of the proceedings. (Or, perhaps, what we thought the proceedings ought to have been. We are mildly disgusted at the mess we made of it by not reading our own documentation, and we have added certain remarks identified as HINDSIGHT which enlarge on this failure.)

WHAT WE HOPED TO GET FROM THE EXERCISE :

From the discussers :

- An appraisal of the suggested method. Does it work ? Could it be improved ?
- An appraisal of the "user interface". Does it constrain the discussion ? Does it enforce uncongenial patterns of argument, thought, or interaction ? Could it be improved ?

- Comments on the acceptability of “formalising” the discussion (by rephrasing the points made to make them easier for the computer to handle and “understand”).

From (or maybe for) the technicians :

- An understanding of the sorts of logical structure which can evolve as the discussion proceeds.
- An assessment of the feasibility of effective “formalising”.
- Suggestions as to other ways in which the computer can help. (Searching for other relevant material, searching for contradictions, etc.)

THE EXPERIMENT.

STEP 0.

INITIAL SYSTEM STATE□

Not in dispute□	Nothing	(□P0□).
Accepted by A□	The health service should be privatised	(□P1□).
Not accepted by A□	Nothing.	
Accepted by B□	Nothing.	
Not accepted by B□	The health service should be privatised.	(□P1□).

HINDSIGHT□If we’d actually begun by writing this down, we might have got further. We would have noticed that there was no common ground between the parties, and perhaps worked out that we needed something there to begin with. This topic will turn up again later.

STEP 1.

Z□ A, please explain why you accept the proposition “The health service should be privatised”.

A□ Because□
Private enterprise encourages efficiency (□P2□; and
Efficiency is an important quality in the health service (□P3□.

SYSTEM STATE□

Not in dispute□	Nothing	(□P0□).
Accepted by A□		
	The health service should be privatised	(□P1□).
	Private enterprise encourages efficiency	(□P2□).
	Efficiency is an important quality in the health service	(□P3□).
Not accepted by A□	Nothing.	
Accepted by B□	Nothing.	
Not accepted by B□	The health service should be privatised	(□P1□).

HINDSIGHT□We immediately observe (□we didn’t, but we should have□) that the lack of an agreed basis (□P0□) has already messed things up. In the previous chapter, we require that A justify P1 by arguing from P0; without that anchor, we immediately fall into the trap□just as we foresaw□of people with “deep convictions which they cannot justify by any rational means”.

NOTE□Have we any way to check the argument step□ In our first try, we left out P3 as a self-evident part of A's position, but without P3 the argument is incomplete. (□In technical terms, we had accepted an enthymeme in place of a syllogism[32]*.□) We noticed that, as in a syllogism, there is a pattern in the argument*[33]*, in that different terms turn up pairwise in the various propositions□ so private enterprise and the health service turn up in P1, private enterprise and efficiency appear in P2, and health service and efficiency are found in P3. That pattern is perhaps something which a computer system can check; alternatively, we can rely on Z to ensure that it's there, but a computer check would be a valuable safeguard.*

NOTE : We put P3 into the argument after noticing that it was missing - but, in practice, people do use incomplete arguments. It is not obvious that we should insist on arguments being completed formally, as we did in our experiments, if only because the formal completion could be very extensive. Perhaps we should accept incomplete arguments if the participants in the debate are satisfied with them. We discuss this matter further in chapter 5.

STEP 2.

Z□ B, please comment on A's chain of reasoning.

B□ Yes, but□
Equal access to health care is more important than efficiency (□P4□).

SYSTEM STATE□
Not in dispute□
Private enterprise encourages efficiency (□P2□).
Efficiency is an important quality in the health service (□P3□).
Accepted by A□ The health service should be privatised (□P1□).
Not accepted by A□ Nothing.
Accepted by B□
Equal access to health care is more important than efficiency (□P4□).
Not accepted by B□ The health service should be privatised (□P1□).

NOTE□ Oops. Something we hadn't foreseen□ the "Yes, but" approach. Perhaps this is uncomfortably common. Certainly, we need to accommodate it, as it is presumably a possibility in any argument based on inconsistent premisses. B accepts A's argument, but attacks its relevance to the point at issue. In this case, the attack is on the grounds of comparative unimportance□perhaps there are other reasons for irrelevance.

NOTE□B has not actually come forward with an alternative proposal; no explicit point of disagreement has been identified. P4 itself is disjoint from anything which has gone before. It's interesting that in this argument we can find a "pseudosyllogistic" pattern of the sort we noticed in the previous step□equal access, important, efficiency, and health appear in P4, efficiency, important, and health appear in P3. We can guess that there should be a P5, perhaps involving equal access and something or other. The obvious possibility, once you've thought of it, is :

Equal access to health care is a more important quality in the health service (□P5□).

It isn't a syllogism, or even an approximation to one, though the pattern is still simple, and certainly one we could implement somehow if we wished to : it follows from the meaning of "more". Even with this step, though, there remains a gap between B's statement and the debate so far. Can we prompt for the missing bits□ For that matter, what are the missing bits□ Are they bits of a meta-argument about what the real argument should be about ? Perhaps they are something like□

> *We must discuss the most important aspect of the health service* (□P6□);
> *therefore*
> *Any argument about efficiency is irrelevant* (□P7□).

This can be seen as an attempt to define more precisely the "should" in P1. There is something mildly plausible about that, but it isn't clear how to fit it into our system. The underlying question is whether we should try to elicit these steps explicitly. If we do (if and when we can), then we may finish up with a much more complete argument, and we could also generate some potential centres of disagreement, which is what we want to do - and we could equally waste a great deal of time simply verifying that the parties to the debate all agree on the details, which they had assumed anyway. We decided that a better course would be to mark the incomplete argument points in some way so that they may be found and inspected later if the need should arise.

NOTE□It seems not unlikely that B, if asked to justify P4, will simply claim that it is an axiom. If A then comes up with an argument intended to refute P4, B might then seek to justify P4 in terms of□for example□a more fundamental belief in equal rights for everybody, having supposed that P4 is such an obvious extension of the equal rights principle as to be unassailable. "High-level" axioms like this can change into arguments.

STEP 3.

Z□ A, please comment on B's position.

A□ Yes, but□
Greater efficiency means that resources are better used (□P8□);
Better used resources will spread further (□P9□);
More widespread resources facilitate equal access (□P10□).

SYSTEM STATE□
  Not in dispute□
    Private enterprise encourages efficiency (□P2□).
    Efficiency is an important quality in the health service (□P3□).
    Equal access to health care is more important than efficiency (□P4□).
  Accepted by A□
    The health service should be privatised (□P1□).
    Greater efficiency means that resources are better used (□P8□);
    Better used resources will spread further (□P9□);
    More widespread resources facilitate equal access (□P10□).
  Not accepted by A□ Nothing.
  Accepted by B□ Nothing.
  Not accepted by B□
    The health service should be privatised (□P1□).

NOTE□ “Yes, but” again. Here, it means something like “I agree, but I didn’t mention it because it’s obvious”. This is rather different from the meaning of the earlier yesbut□ now A is claiming B’s argument in support of his own. The issue is blurred, and it now appears that private enterprise leads to both efficiency and equal access. B can only make progress by attacking some step of the argument.

NOTE□ For what it’s worth, the argument advanced by A is a polysyllogism[34]*; it is simply a chain of syllogisms with the successive conclusions left unstated.*

STEP 4.

Z□ B, please comment on A’s chain of reasoning.

B□ I contest A’s argument linking efficiency with equal access. It is a historical fact that□
Increased efficiency leads to unequal access (□P11□).

SYSTEM STATE□
Not in dispute□
Private enterprise encourages efficiency (□P2□).
Efficiency is an important quality in the health service (□P3□).
Equal access to health care is more important than efficiency (□P4□).
Accepted by A□
The health service should be privatised (□P1□).
Greater efficiency means that resources are better used (□P8□);
Better used resources will spread further (□P9□);
More widespread resources facilitate equal access (□P10□).
Not accepted by A□ Nothing.
Accepted by B□
Increased efficiency leads to unequal access (□P11□).
Not accepted by B□
The health service should be privatised (□P1□).
* Greater efficiency means that resources are better used (□P8□);
* Better used resources will spread further (□P9□);
* More widespread resources facilitate equal access (□P10□).

*NOTE□ B has simply denied the validity of A’s argument by appealing to experimental observation. This is legitimate, but doesn’t get us very far. It shows (□if valid□) that at least one of P8, P9, and P10 is invalid, but not which one. (□These propositions are marked with * in B’s list of unaccepted propositions in the state specification.□) B can reasonably be challenged to produce the experimental evidence, and cannot reasonably insist that A produce evidence to the contrary.*

NOTE : The introduction of experimental evidence shows that we should make provision for the system to remember the reason for believing (or disbelieving) each assertion it records. Some assertions are supported by evidence, some are supported by argument, some are supported by authority, some (the axioms) are not really supported at all. The nature of the support is important in seeking for weak points in an argument. Good experimental evidence is unassailable, while axioms are fair game.

NOTE: An alternative approach for B would be to present an argument:

B: I contest A’s argument linking efficiency with equal access.
Increased efficiency inevitably means reduced staff time (P12).
Reduced staff time means rushed consultations (P13).
Rushed consultations give worse service to unassertive people (P14)

SYSTEM STATE:
Not in dispute:
Private enterprise encourages efficiency (P2).
Efficiency is an important quality in the health service (P3).
Equal access to health care is more important than efficiency (P4).
Accepted by A:
The health service should be privatised (P1).
Greater efficiency means that resources are better used (P8);
Better used resources will spread further (P9);
More widespread resources facilitate equal access (P10).
Not accepted by A: Nothing.
Accepted by B:
Increased efficiency inevitably means reduced staff time (P12).
Reduced staff time means rushed consultations (P13).
Rushed consultations give worse service to unassertive people (P14).
Not accepted by B:
The health service should be privatised (P1).
** Greater efficiency means that resources are better used (P8);*
** Better used resources will spread further (P9);*
** More widespread resources facilitate equal access (P10).*

Here again B tackles the compound assertion put forward by A without unambiguously invalidating any of its components.

NOTE: B has not yet said anything about state control. A has had the initiative throughout simply because A was the first to be asked to present an argument. It is true that we have to consider A’s arguments sometime, so perhaps it doesn’t matter who starts off; in practice, though, a long delay before putting one’s point of view can be very discouraging.

That’s as far as we got with the “debate”. We then spent some time on discussing the whole thing.

WHAT WE GOT OUT OF IT.

Before the account of the experiment, we listed a number of results which we hoped to gain. We review these hopes here.

From the discussers, we expected :

- **An appraisal of the suggested method. Does it work ? Could it be improved□**

We observe first, from the HINDSIGHT of step 0, that the method would probably have been more effective if we had used it ! On the other hand, the variety of topics introduced in the argument as it proceeded seems to show that no simple agreed base of statements will be logically adequate, short of a complete implementation of general knowledge. (Which is, of course, the impetus behind the Cyc work[31].) Perhaps that should not have been a surprise. If it were really easy to write down a complete set of mutually acceptable axioms, someone would have done so, and at least some of the festering questions might have been settled long ago.

> (Then again, they might not : people have great facility in ignoring uncomfortable facts. In the last few weeks - writing on 23 March 1991 - there has been much discussion in the press of the case of a Samoan convicted of serious crimes, who has escaped deportation by applying for, and being granted, New Zealand citizenship while in prison. There are, predictably, howls of outrage - and I have seen no mention at all of the reasons for the Samoans' special position with regard to New Zealand citizenship, canvassed in great detail in the same newspapers only a few years ago.)

Recalling that we expect to find our results in the areas of disagreement between the participants, it is a little unsettling to find that even after step 4 we have found nothing that A does not accept. This is perhaps partly because the initiative has so far been entirely with B, and clearly there is a lot further to go in the debate. Only experience will show whether this is in fact a serious problem.

We finally noticed the lack of W. The result was incoherence□ anyone can at any time introduce a new approach into the discussion, without having to justify it in any way. In default of an actual treaty or other agreed document, we thought it would be useful to provide facilities for each participant to begin by presenting a position paper, which would be a summary of axioms and principles assumed by that participant from the outset. The collection of universally accepted items from the participants' position papers would stand in place of W; the remaining items would form the initial agenda for discussion. Facilities to help people construct their own position papers could be useful. These could help to ensure that a position paper was internally consistent, but would not involve any interaction between different people's papers.

What do we do if the position papers show no common ground ? We are then back in the position of having no W, which seems to lead only to confusion. Can we somehow help the participants to find some principle on which they can agree ? Any such principle must, of course, be relevant to the question at issue : agreeing that the sky is sometimes blue is unlikely to be constructive. A useful common principle must also be sufficiently powerful to act as a satisfactory starting point for all parties, so the search must start from the position papers and work towards more fundamental beliefs until some point of agreement is found. We do not explore this possibility further here; and we do not know what to do if no common principle appears. We do point out that this discussion emphasises the central importance of the Treaty of Waitangi in our scheme.

- **An appraisal of the "user interface". Does it constrain the discussion ? Does it enforce uncongenial patterns of argument, thought, or interaction ? Could it be improved ?**

  B's contribution in step 4 introduces an element of uncertainty into what we have defined as the system state : there are some statements which we cannot classify as either accepted or rejected by B. We could add a new category of don't-know statements (where it is the system which doesn't know, not B); but that would be quite hard to administer adequately. In this example, after step 4 it would be the set { P8, P9, P10 }. Now, suppose by some means we find that B accepts P8 and P9; then (assuming that B is arguing logically) we may presumably safely assume that B rejects P10. If we find that B accepts P8 and rejects P10, though, we can say nothing about B's stand on P9. (And there is always the possibility that in fact B accepts all the three statements, but denies the validity of the argument put forward by A in step 3.)

  This is rather unsatisfactory. We can get out of it by asking B immediately to be more precise about the defects of A's statement, but B may not be able to do so. We can simply forget about the disputed statements, at least so far as the state is concerned; or we can write our rules so that Z quizzes B about such statements as a matter of priority. At the moment, we have no satisfactory solution to this problem.

- **Comments on the acceptability of "formalising" the discussion (by rephrasing the points made to make them easier for the computer to handle and "understand").**

  We remarked that the pattern was reminiscent of those characteristic of syllogisms. Perhaps we want to insist wherever possible that each step really is a syllogism[35]. That gives a much better chance for computer checking, and also supplies a clear structure for the argument step. On the other hand, this could be seen as a constraint. Should people be permitted to get away with bad logic in the interests of freedom□ One can argue that they should not□the intention is to come to some sort of reliable conclusion, not to let people play silly tricks on each other. There would be much more point in restricting the argument to syllogisms if that solved any problems - but, unfortunately, it doesn't. It remains true that a major loophole is the essential incomprehensibility (to the computer) of the terms of the argument themselves, and until we can do something about that, other precautionary measures are probably no more than patches. An example is the problem of "should" mentioned earlier.

From the technicians, we expected :

- **An understanding of the sorts of logical structure which can evolve as the discussion proceeds.**

  The big surprise was that arguments could not only be rebutted; they could also be yesbutted. Perhaps it should not have been a surprise - after all, we do it all the time - but it was unexpected, and in consequence our system was unready for it. In logical terms, the yesbut argument breaks out of the pattern of regular development of a tree structure to start a new tree, or maybe a new branch of the existing tree. In consequence, the structure which describes the argument is not a tall, narrow tree, but something much more like a broad, squat forest.

  It is important that this unexpected argument structure does not alter any of the principles on which we have based our thoughts. We are still aiming for the same target : acceptable logical arguments describing the positions of each of the participants, from which we can identify the fundamental disagreements which go to form the schemata which underlie their approaches to the Treaty of Waitangi.

Nevertheless, it may be that we could handle the actual arguments more effectively using different sorts of structure. We do not yet know if this is so, and only further experience will clarify the issue. As an example of a possible line of development, if it seems that the broadness of the "forest" makes it difficult to correlate different lines of argument, it could be worth seeking structures which emphasise the connections between different threads.

- **An assessment of the feasibility of effective "formalising".**

The retreat from formal logical structure does not necessarily mean that we have no structure left - though it may be that such structure as remains is weaker. We saw one example in the discussion on step 1 : even though we may no longer express our arguments as formal syllogisms, the characteristic pattern of three terms may still remain, and is commonly evident even at a purely textual level. This is something we can seek and record, and if we don't find it we may seek clarification. Without a good level of natural language understanding by the computer, this understanding will of necessity be supplied by people.

It may be possible to direct some purely formal queries to administrators rather than to participants. An apparently missing term may turn out to be merely a matter of rephrasing, or a missing component may be general knowledge. It would be unwise to burden the participants in the debate with comparatively trivial questions. There is a danger in this technique, though : if we rely on an administrator's general knowledge, we risk importing the administrator's own schema into the debate, and we shall not notice it. Provided that the computer's part in the argument remains subsidiary, this may not matter much, but it is certainly a factor which should be borne in mind.

It might be helpful if the computer system could be involved in overseeing the progress of the debate, perhaps by checking that steps are performed in sensible order, and making sure that required information is collected. For example, at any time in a debate there may be many ways of proceeding, each of which is in terms of logic equally valid. Not all of these ways may be equally effective, though; in a well conducted argument, each step carries on the discussion in some planned direction, and unless the direction is more or less maintained, the argument gets lost in a maze of logical but unhelpful comments, remarks, interjections, and other incidentals. Can we use the computer to keep the debate on track ? To do so, we need to know just how each statement contributes to the argument; it is unlikely that the function of an argument step can be inferred from its content, so⊟if we want it⊟this is an item which must be provided or elicited by Z. A reminder would help.

- **Suggestions as to other ways in which the computer can help. (Searching for other relevant material, searching for contradictions, etc.)**

Some suggestions as to further computer activities can be drawn from the preceding discussion in this section. Perhaps the most valuable would be to insist on proper definition of the basic components of the state before beginning. While it is likely that the forest would have grown by yesbuttal in any case, it could perhaps have been confined had there been a recognised basis of principle from which to argue.

It may be, of course, that a forest would do, if only we could work out what it meant. Whether it grows by yesbuttery or by loosely directed discussion, there may be enough solid wood in the statements made to reach some sort of conclusion. Any help in classifying, sorting, or otherwise collating an amorphous set of assertions could be valuable in sorting the pearls from the swine. Or, for that matter, the silk purses from the sows' ears.

STRUCTURES.

We had hoped that the experiment would throw some light on the sorts of data structure which our system would be required to handle. There appear to be two important considerations for which any structural design must provide□the arguments advanced, and the people involved. Unfortunately, it is far from clear how these can be separated. It is easy enough simply to list the statements made, and to connect these together into steps in the argument, and also to list the people. It is also easy to record who said what. All these are matters of readily accessible fact.

The whole point of the exercise, though, mixes together the two levels in a rather messy way. To resolve the argument, we must find out who believes what and why, and allow for people to change their beliefs as the debate progresses. For example, we could be required to provide mechanisms which can record that A believes that the moon is made of green cheese, on the grounds that everything turns green at night, and everything in the sky is made of cheese, while B believes that the moon is made of rock on the basis of the United States NASA experimental findings. This gives us two contradictory statements to begin with. (□We point out that the "readily accessible facts" mentioned above are that participant A advanced proposition P, not that any proposition P is itself factually true.) Eventually, after a long and complex argument, B succeds in winning A to the NASA point of view; whereupon we must somehow record that A no longer believes in the green cheese hypothesis, and, presumably, do something about any conclusions we have stored which depend on that hypothesis - and also amend A's premisses, which still lead to the green cheese conclusion. This gets us squarely into the area of truth maintenance systems[36].

From the point of view of data structures, it is unclear whether the argument part of the structure should be seen primarily in terms of attributes of the assertions, or of the people, or as a separate third group of structures. We propose the two-level structure sketched below as a first guess; we expect that further experience will clarify the issue, and lead to greatly improved structures. First, there is the substance of the arguments themselves : the assertions, the connections between them, and the conclusions which are drawn. (□which are, of course, simply further assertions). At this level we make no commitment as to the validity of the assertions; we cannot, as this may be a matter of dispute.

The **assertion** is the basic data type. It is composed primarily of a text string, which is the text of the assertion itself. In addition, as we observed in discussing step 4, we need some information on the status of the assertion : is it supported by experimental evidence, or argument, or is it an axiom ? A complication is that there may be disagreement over the status : what A accepts as an axiom, B may reject by argument. It is nevertheless true that both A and B have made assertions, so they should be recorded. A list of arguments which support the assertion is also needed, and a complementary list of argument steps in which the assertion is used would help. particularly if the assertion should be withdrawn. Perhaps there should also be a list of arguments which tend to oppose the assertion.

An **argument step** corresponds (□reasonably enough□) to one step of the argument. It is something like a frame structure, containing the conclusion of the step, a list of the statements used in reaching the conclusion, and who used the step. It may also contain some identification of the argument in which it appears, and some indication of its function within the argument□ does it support some other statement, or attack some statement□

An **argument** is a list of argument steps. This might, in fact, be a better place to keep the functions of the argument steps, because they have meaning only with respect to the argument as a whole, and not within the context of a single argument step.

Second, there is the place of these arguments in the debate. Who said what, and in what context ? How is it related to other statements made by the same, or other, participants ?

The **participant** is a person who takes part in the debate. The structure representing a participant must include the participant's current sets of accepted and not accepted assertions - and, perhaps, the participant's current sets of accepted and not accepted argument steps ?

The **contribution** of a participant is a chronological list of assertions made by the person. (This would normally be composed of pointers to the assertions in the arguments.)

Whatever structures we choose, it's important to be able to save the current state of the system and restart it. People will want time off to think, and to consider their responses.

A PARTIAL IMPLEMENTATION.

Some of these ideas have been embodied in an experimental implementation[37]. While far from complete, it illustrates some of the points discussed in this chapter. In particular, it contains structures representing propositions, arguments, participants, and the system state. It differs from the organisation proposed here in several significant respects :

- The argument steps have been artificially simplified. They must be presented explicitly in the form

  <proposition 1> ▯ □<proposition 2>

  and, as that form implies, they are restricted to include only a single premiss.

- There is no explicit structure to represent an argument step. This is partly because, with the steps restricted to the simple form described, an argument step can always be represented by a link between two propositions. Each proposition therefore contains a list of all the propositions it supports, and another of the propositions which it negates.

- There is no explicit representation of an argument either. While all the steps are recorded, the sequence is lost, so it is not possible to determine the context in which each point is made. This is not an approximation in strict logical terms, as the full set of assertions is retained, but it could be unsatisfactory is it were desired to retrace an argument.

- There is no record of who advanced each argument step; it is supposed that knowing who put forward each proposition is sufficient. This is again not a logical flaw, but as different people may base different conclusions on the same premiss, some information on who thinks what is lost, and that could affect the construction of the schemata.

- The programme imposes a cyclic order of contribution on the participants. While it is always possible to make that work (by allowing participants to pass on any cycle), it is perhaps too formal to be a satisfactory base for a free discussion.

Despite these defects, the exercise was instructive, but it is clearly necessary to extend the representations used to record the argument if the implementation is to be useful in practice.

CHAPTER 5

ASSESSING OUR WORK.

What have we achieved in this project ? We have certainly not achieved the computer implementation of the discussion moderator which we envisaged when we began. (It may be worth remarking that the original proposal was for an M.Sc. thesis, entailing one student's full-time work for a year. In the event, Roy Davies took it on as a "project", involving one sixth of his time for about 8 months.) We have achieved a number of more or less intangible ends : we have a much clearer idea of what is involved in the task we set out to accomplish, both in terms of the structure of the task itself and in the nature of the computing techniques which will be needed to make further progress.

SCHEMATA.

Since our initial first flush of enthusiasm sparked off by Bartlett's work with American aborigines, We have wondered a bit about the idea of group schemata. Introspection has suggested to us that, if groups can be identified as collections of people sharing a common schema, then everyone is in a group of one. As we see no particular reason to believe that we're special, then perhaps everyone else also forms a group of one. So much for group schemata.

That is not to deny the value of the psychological work, but rather to suggest that it deals with something which isn't quite the same as that with which we expect to deal in our computer system. The psychological studies were, we suspect, essentially studies on groups, and report characteristics of group activity; we have always thought of the computer project as involving individuals. Doubtless you could fit groups in by putting every question to discussion in the group before presenting an agreed answer to the system - but, as the system itself is supposed to be able to cope with discussions, that does seem to be something of a roundabout way of doing things.

We would therefore expect to end up with a complex set of individuals' beliefs and differences, which may or may not fall into clusters of broadly similar patterns, identifiable as the group schemata found in the psychological studies. Maybe there *are* schemata; but if so we should be able to identify them in the structures produced by our system. Our earlier attempt to build them in from the beginning was perhaps unwise.

That doesn't mean that the idea of schemata is dead, though, and there are two ways in which our work (should it ever be completed) could contribute to identifying something in the nature of schemata. First, we saw the possibility (in the imaginary experiment of Chapter 3) that a participant may base an argument on axioms. Some of these axioms may well be part of the participant's schema for the topic discussed. Just how to decide which axioms are part of a schema and which aren't needs some distinguishing criterion, but the possibility is there. On the other hand, only experience will tell how common such argument directly based on schemata might be; the very nature of schemata suggests that they are more likely to remain hidden than to appear overtly. They are the *unconscious* assumptions rather than those which can consciously be marshalled in logical arguments. They are nevertheless still present in the arguments, and the second remark is about the possibility of identifying them automatically. This suggestion is - obviously enough - speculative to a high degree; but some of the machinery needed appears to be available. Constant, Matwin, and Oppacher[38] describe a technique which they call LEW (LEarning by Watching - clearly, the authors were not brought up in a cricketing country) whereby an "intelligent" programme is able to infer reasons for answers to questions from analysis of the dialogue concerned. It is notable, though, that LEW requires that the dialogue be represented in structured form, as its basis for operation is comparison of structures. In this it resembles the linguistics of Broomfield and Harris. (We have been here before. There is little if any difference in principle between the "standard form" required by LEW and Schank's "interlingua".)

While more study would be needed adequately to evaluate LEW as a potential component of our system, it seems at first sight that our information structures match its requirements very well. For raw material, we may use either (or perhaps both) the assertions or the arguments; in both cases we have enough structure to satisfy LEW's requirements. All that is needed is sufficient detail to support partial matching between different components - so, given P's and Q's statements about sheep, LEW could transfer its knowledge about sheep to other nouns by making appropriate matches between words. (In this, it can be thought of as behaving as a learning version of Eliza.) It could perhaps perform similarly at a higher level given the structure of argument steps as data. In all, it seems likely that LEW's techniques would repay further investigation.

REFORMULATION.

The combination of the misgivings expressed in Chapter 2 and the observation that people do not in fact conduct Platonic ideal arguments gave us cause to reconsider some of our ideas. We became convinced that we could not afford even to rephrase people's statements in what we earlier called "more formal" terms, at least so far as the primary arguments are concerned. The risk of twisting meanings is too great, particularly as different people may perceive different twists in the same "formalisation". A possible consequence is that much energy would be diverted from elaborating the real argument into subsidiary, and probably unproductive, arguments about the correctness of the formal statements.

We therefore respecify the aim of this study :

> *We hope to develop a discussion moderator, which will keep track of chains of reasoning in a discussion, identify and try to resolve differences in points of view, and compose collections of points of agreement and disagreement between the parties concerned.*

The aim is to record what was said, and what the participants *thought* they meant, and make it available for reference and further questioning. What's happened to the schemata ? They've been absorbed into the "points of agreement and disagreement", where they are distinguished from other points of agreement and disagreement by being asserted as axioms which are not universally accepted.

There is no mystery about recording "what the participants thought they meant"; we do not require empathic software, nor telepathic hardware. Participants contribute to the debate by offering observations of two sorts : simple facts, and argument steps. If I say "Daisy is a cow", then "what I think I mean" is simply the text of the statement. If I say "Daisy is a cow, and all cows eat grass, therefore Daisy eats grass", then "what I think I mean" is that the conclusion of the statement follows from the premisses. What is recorded is therefore a set of text strings, and a set of structures representing argument steps which link the strings together.

Another definition : the phrase "the conclusion of the statement follows from the premisses" is not to be interpreted as a strictly logical condition. A proper logical statement is, of course, acceptable, but many argument steps as given are, from a standpoint of precise logic, incomplete, because they assume a lot of general knowledge or a cultural background which it would be tedious or impossible to define in any precise sense : "Like all cows, Daisy eats grass". The phraseology in the previous paragraph is chosen with some care : "I mean that the conclusion of the statement follows from the premisses". Other participants may disagree - which would hardly be possible if we insisted on pure logic. On the other hand, if all parties accept the statement, then whether or not it is strictly logical is immaterial.

There is one question to ask about that convention. When I put forward the argument step, should that be taken as asserting the text of each of its premisses, as if I had offered them as simple facts, or should the argument step be assumed as a hypothetical proposition, valid whether or not its premisses are accepted ? (In other words, should we interpret the step in the example as "IF Daisy is a cow, and IF all cows eat grass, THEN Daisy eats grass" ?) It seems clear that only the conditional interpretation is practicable; if we took the other view, we might be unable to accept the argument step if we already knew an earthworm called Daisy. Nevertheless, it is not unusual to put forward argument steps without separate justification of the premisses; should the system check for support for premisses, and ask for it if it is not forthcoming ?

We have assumed so far that contributions to the debate always have conclusions, but that isn't necessarily so. I could replace my assertions about Daisy by the remark "Of course, all cows eat grass" - leaving it to the other participants to apply this general statement to Daisy, a perfectly reasonable assumption in a debate between people, but far from easy to implement by computer means. Does this affect our system design ? We believe that it doesn't. We are still trying, primarily, to record what was said; in such an abbreviated argument step, we are less able to determine what the arguer means, but we can still regard the statement as an assertion of fact - which, of course, it is. We came across an example of this in the "real experiment" in discussing B's assertion that "Equal access to health care is more important than efficiency".

This is not to suggest that the attempt at formal representation is itself a waste of time, though. It may well even be an essential part of the computer side of the system. If our system is to be an effective means of recording and checking the arguments, it must have some way of carrying out the checks; and if it is to go beyond this function, and seek discrepancies and contradictions between assertions perhaps made at different times and in different contexts, it must have enough information about the structure of the arguments in a form which it can handle to perform the required computations. This sort of computation is likely to become more and more important as the body of knowledge available to the system increases. We would hope that the computer system would identify possible conflicts, and raise questions about them; but these would be regarded as queries only, not as indisputable parts of the argument.

WHITHER FORMAL REPRESENTATIONS ?

The reformulation described takes away from the system any direct involvement with the logic of the argument; it becomes an uncritical administrator, keeping track of the flow of debate and recording people's pronouncements reliably for future reference. Must we, then, abandon our hope of using computer techniques to check the argument itself ?

In the short run, perhaps we must - and perhaps this limitation is sensible, in the interests of getting a working programme. In the longer term, though, the answer could be no, for two reasons.

- First, some of the pronouncements will be in logical forms which we can interpret without fear of error. If P should say, "I assert that all sheep are white, because all my sheep are white", then we can safely record something of the form :

  P says : all-my-sheep-are-white □ all-sheep-are-white.

  Here, we encode only the purely logical structure of the assertion. What we cannot do is undertake to encode the component statements in some logical form with which the machine can work; in this example it would be quite easy, but in general it would not. The system will therefore not be able to detect the faulty implication in the assertion (or, alternatively, infer that P owns all the sheep there are). It will, though, be able to identify the contradiction following from a subsequent assertion encoded as :

  Q says : I-have-a-black-sheep □ not all-sheep-are-white.

  True, we rely on human intelligence to perform the encoding, but the possibility is there; and if people are debating a set of statements it is not unlikely that some will be directly challenged in a form which lends itself to such encoding. Much depends on the development of a suitable vocabulary in which to conduct the argument.

- The second reason to hope that machine logic will be useful is the possibility of achieving some, however limited, mechanical interpretation of language. Provided that this level of activity is seen as advisory only, and kept separate from the real arguments, all should be well. The interpretation does not need to be highly sophisticated; even rephrasings after the pattern of Eliza[39] could perform conversions such as :

  I have a black sheep □ not all sheep are white,

  though how to prevent the generation of vast numbers of valid but useless assertions (I own things, I own a sheep, some sheep belongs to me ...) could be an interesting problem.

REORGANISING THE SYSTEM.

Our major aim remains unchanged, but the experienced recorded in this report has certainly taught us that our initial emphasis on logic was a mistake, but that there is an equally taxing, and at least equally important, clerical task to be performed. What are the implications of these reflections on our grand plan ? To put it another way, if we were now to begin again, what would we see as the important goals to be attained and techniques to be used ? We address these points in the following sections.

CLERICAL.

In most respects, this new view of the process we wish to administer makes rather little difference to any but fine details of the way in which we should do it. The source of all the activity is still a debate between two (or, in principle, more) parties; we still need to record the points made; we still need to know who said what, and when. On the other hand, we no longer expect that each step in the argument will look like an example from a textbook in elementary logic. In some ways, this simplifies the requirements : all we need save is the original text of whatever was said, and no internal structure need be defined. The move away from formality also makes it harder to organise the contributions to the debate for information retrieval purposes - at the same time as it makes the ability to look at the original text more important. Any question about an argument step can now only be resolved by inspecting the associated text, and perhaps reinterpreting it. Provision for full-text keyword search and retrieval would help to find specific argument steps, and also other statements which may relate to a topic under discussion, and some such provision becomes more important under the new regime.

A system described as a “personal information manager”[40] has specifications not unlike those which we require : to deal with unconstrained text, to extend its structure to accommodate new text, and to express in its structure certain relationships between the items entered. There are differences too, but the common ground suggests that we might be able to profit from studying their system.

LOGIC.

Previously, we had some chance of using the formal structure as a basis for checking the argument. This structure is no longer available. If we still wish to exploit the computer’s ability to check logic, we must find another way. Two ways suggest themselves.

First, we can revive the idea of natural language interpretation. This hasn’t got much easier since we started work on the project, but circumstances have changed. Now we are no longer supposing that the whole debate should depend on the programme being able to understand natural language, as the real thinking is (□very properly) left to the people. It may well be possible to get enough sense out of the text to form a rough idea of what was said, and to check on that basis; the checking is now advisory rather than central to the operation of the system, and not guaranteed - but it may be a lot better than nothing at all, and has a fair chance of some success so long as people conduct their arguments in comparatively simple language.

Second, we can continue to ask the participants to provide formal versions of their statements. We would still regard the original text as the primary argument, but the formal version would be a satisfactory basis for checking, indexing, and searching. This should be much more precise and reliable than the first approach.

We have not so far implemented either of these checking methods, but we lean towards the first. Its main advantage is that it is quite automatic; the participants do not have to worry about reformulating their statements, which would certainly be an advantage if the system were ever used by people without experience of formal argument. A possible defect in the second technique is the possibility of unbalance between the Maori and English natural language systems. This possibility is unavoidable; but recall that it does not affect the primary argument, only the backup system, and that if the argument is well conducted it should not be needed at all.

Another reason for preferring the natural language approach is that it gives an opportunity to explore a more effective dual language system. It would be instructive to see how far, and how adequately, we can in fact handle the subtleties and nuances which we expect to make natural language work difficult, and to what extent we can use a common “interlingua” to represent the texts.

THE NEXT STEP.

Have we enough experience to decide which way to go from here ? We believe that we have, and we’re going to try anyway. To do so, we must identify clearly the end which we hope to achieve, the data structures which we shall need to represent the information acquired during a debate, and the operations which we wish to perform on the information.

The aim.

Our aim is still much as we redefined it in Chapter 3; but we might like to interpret this in the light of our second experiment, the implementation, and the previous discussion in this chapter. Implications of the reinterpretation will appear as we proceed, but their general tenor is to extend and strengthen the means of access to the debate, both to facilitate largely manual operations now and to prepare for more automatic help in the future.

The data structures.

The data structures identified in Chapter 4 still seem to be appropriate.

The question of where to record the function of a step within an argument is complicated a little by the appearance of yesbuttery; but this can be seen as reinforcing the notion that the function should be kept at the argument level, not as an attribute of one argument step.

The possible resurgence of formal representations when we find out how to do it is probably better handled through a separate parallel set of structures rather than by including formal equivalents in the parent structure; we saw that there could be disagreement between participants as to the appropriateness of a proposed formal equivalent.

Similarly, at the assertion level, we may wish to include connections between equivalent assertions in Maori and English - and, once again, there is room for disagreement as to the validity of the proposed equivalent.

The nature of the links in the two examples just discussed is not simple, as is clear from the potential for disagreement. Indeed, as we saw in Chapter 3, the assertion that two assertions are equivalent should be treated in the same way as any other sort of assertion in the system; its representation may therefore be complicated.

The operations.

We have not previously worried about the operations we wish to perform on the structures. This has not, of course, cramped our style in actually performing the operations as and when it seemed convenient; but now we should review what we have done, and try to isolate the components which we believe will be useful in future.

The basic operation is to add a step to an argument. This involves not only recording the argument step with its own attribute of who said it, but also constructing links with agreeing and dissenting opinions, augmenting the current argument structure to include the new argument step and its function in the argument, and making appropriate changes to the information about the participants, particularly in adjusting their lists of accepted and rejected statements.

We haven't used any so far, but we shall presumably eventually need facilities to withdraw assertions and argument steps. When this happens, it will be important to follow through the consequences of the withdrawal on arguments known to the system.

Other immediately useful operations are full-text keyword searches to find assertions relevant to topics at issue, and access to all aspects of the system state - who believes what, and who doesn't, what the current disagreements are, and so on.

ONWARD (OVER THE SEA ?) TO SCAE.

We introduced Orts and Scae in Chapter 1 as two supposed "components" of Socrates. Orts is the half we have been discussing throughout this report; it administers a debate, but understands nothing that transpires and, therefore, can contribute nothing of its own. Scae is the other half of Socrates, which not only comprehends the debate but also joins in. (And, if Plato is to be believed, always wins.) Do our investigations lead us any closer to an implementation of Scae ?

So far as we can see, the quick answer is that Scae is just as far off as ever. A successful implementation of the system we have described would provide a great deal of structure to support the material of the debate, but the nature of that structure is determined by the participants in the debate, and there is no sense in which the structure, or any of the assertions which are its raw material, is *understood* by the system. A system which takes an active part in an argument has to do a lot more than that. Specifically, it must in some sense understand the material it handles. In practice, that doesn't turn out to be a very useful statement, because it serves more as a definition of what we mean by "understand" in this context, but it does emphasise the point that there are substantial activities which Orts doesn't have, but which Scae will need, and that they have to do with probing into the sentences which Orts stores and retrieves.

But there is also a slow answer. Work of the sort introduced briefly in this chapter, like the conceptual dependency techniques described earlier, offers a way to probe more deeply into the material of the statements made in the debate. Even if this does not constitute understanding in any human sense, it may lead to a sufficiently reliable analysis to reveal common material in different statements, and this can be used as a basis for making suggestions about the statements which the system can put to the participants. Even a suggestion of the form :

Is there not a conflict between P's assertion that "all sheep are white" and Q's observation that "I have a black sheep" ?

could be a significant contribution, and appears to be well within the capabilities of Schank's work. (It also has something of a Socratic feel about it, which may or may not be significant.)

A possible alternative approach is that of LeClair[41], who describes a knowledge acquisition system which can accept knowledge from several experts. The aim is to combine knowledge from the several experts in order to reach a satisfactory solution to a problem, and also to use the several experts' different approaches to derive new knowledge which can be used in future. An assumption which fits well with our aims is that the experts are competitive rather than cooperative. LeClair says, "The key to accommodating multiple, potentially conflicting experts is to manage their interaction while maintaining knowledge base consistency". He appears to use "consistency" to mean consistency *within the current argument*, but not necessarily to include implicit contradictions between different experts' knowledge; for consistency is maintained by accepting as irrefutable any information specifically relating to the current problem, while ignoring any conflicts between different experts' intermediate conclusions. Our interest is to *identify* the conflicts between the "experts'" knowledge which leads to the disagreement. Nevertheless, there are clearly parallels between LeClair's work and our own, and again it might be possible to extract some suitable Socratic suggestion which could point to unrecognised conflicts in the argument, and perhaps lead to their resolution.

CONCLUSION.

We do not adopt the traditional computist's position that our work is 95% complete - we would be brave were we to insist that what we have achieved would amount to the other 5% - but we believe we have made a start on an interesting and worthwhile field of investigation. Should our aim of producing an effective debate moderator be achieved, we would have a significantly new sort of computer programme; even if it isn't, we have found some intriguing and stimulating problems on the way.

REFERENCES.

1 : Reference 42, page 13.

2 : B. Russell : “Mathematics and the metaphysicians”, an essay reprinted in *Mysticism and logic* (□Penguin book reprint, 1953), page 79.

3 : G.A. Creak : *An application for schemata*, unpublished Working Note AC70, 1989.

4 : G.A. Creak : *Schemata and the Treaty of Waitangi*, unpublished Working Note AC74, 1989.

5 : N.E. Sharkey, G.D.A. Brown : “Why artificial intelligence needs an empirical foundation”, Chapter 11 of reference 43.

6□ J. Ramirez G. : Use of structure-based models in the development of expert systems, *Sigart Newsletter* **#109**, 35 (□July 1989□).

7 : Reference 42, page 164.

8 : Plato : *The Republic* (translated by F.M. Cornford, Oxford University Press, 1941) page 30.

9□ W. Lehnert□*The process of question answering* (□Lawrence Erlbaum Associates, 1978□).

10□ Reference 9, page 17.

11□ Reference 9, page 108.

12□ R.C. Schank□*Conceptual information processing* (□North-Holland Publishing Company, 1975□).

13□ Reference 9, page 248.

14□ Reference 12, page 8.

15□ Reference 12, page 30.

16□ Reference 12, page 23.

17□ C.J. Rieger□“Conceptual memory and inference”, in Reference 12, page 170.

18□ Reference 9, page 222.

19□ Reference 9, page 251.

20□ Reference 12, page 168.

21□ R.J. Brachman, J.G. Schmolze□“An overview of the KL-ONE knowledge representation system”, *Cognitive Science* **9**, 171 (□1985□), reproduced in reference 44, page 207.

22□ R.J. Brachman, V.P. Gilbert, H.J. Levesque□“An essential hybrid reasoning system□knowledge and symbol level accounts of Krypton”, *IJCAI-85*, 532 (□1985□), reproduced in reference 44, page 293.

23□ S. Jones, P. Mason, R.K. Stamper □“Legol-2□a relational specification language for complex rules”, *Information systems* **4#4** (□December 1979□), reprinted as *L.S.E. papers in informatics #L27* (□London School of Economics□).

24□ R.K. Stamper□“Legol□modelling legal rules by computer”, from B. Niblett (□ed□)□*Computer logic and legal language* (□Cambridge UP, 1980□), reprinted as *L.S.E. papers in informatics #L35* (□London School of Economics□).

25□ M.J. Sergot, F. Sadri, R.A. Kowalski, F. Kriwaczek, P. Hammond, H.T. Cory□“The British nationality act as a logic program”, *Communications of the ACM* **29**, 370 (□1986□).

26: P. Leith□“Fundamental errors in legal logic programming”, *Computer Journal* **29**, 545 (□1986□).

27□ R. Kowalski, M. Sergot□“Leith and legal logic programming”, *Computer Journal* **30**, 285 (□1987□).

28□ D.H. Berman, C.D. Hafner□“The potential of artificial intelligence to help solve the crisis in our legal system”, *Communications of the ACM* **32**, 928 (□1989□).

29 : G. Orwell : *Nineteen Eighty Four* (Penguin Books, 1954) : particularly the Appendix on “Newspeak”.

30 : Reference 29, page 249.

31 : D.B. Lenat, R.V. Guha, K. Pittman, D. Pratt, M. Shepherd : “CYC : towards programs with common sense”, *Communications of the ACM* **33#8**, 30 (August 1990).

32□ Reference 45, page 71.

33□ Reference 45, page 54.

34□ Reference 45, page 70.

35□ Reference 45, page 53.

36 : E. Charniak, D. McDermott : *Introduction to artificial intelligence* (Addison-Wesley, 1985), page 414 (where it’s called “reason maintenance”).

37 : R. Davies : *The use of a computer in discussion mediation, a feasibility report*, Project report, Auckland University Computer Science Department, 1990.

38 : P. Constant, S. Matwin, F. Oppacher : “LEW : learning by watching”, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 294 (1990).

39 : J. Weizenbaum : “ELIZA - a computer program for the study of natural language communication between man and machine”, *Communications of the ACM* **9**,36 (1966).

40 : S.J. Kaplan, M.D. Kapor, E.J. Belove, R.A. Landsman, T.R. Drake : “Agenda : a personal information manager”, *Communications of the ACM* **33#7**, 105 (July 1990).

41 : S.R. LeClair : “Interactive learning : a multiexpert paradigm for acquiring new knowledge”, *Sigart Newsletter* **#108**, 34 (April 1989).

42□ R.H. Thouless□ *Straight and crooked thinking* (□Pan Books edition, 1953□).

43 : M. Yazdani (ed) : *Artificial intelligence : principles and applications* (Chapman and Hall, 1986).

44□ J. Mylopoulos, M.L. Brodie (□ed□)□ *Readings in artificial intelligence and databases* (□Morgan Kaufmann, 1989□).

45□ L.S. Stebbing, *A modern elementary logic* (□Methuen, 5th Edition, 1957□).